PL-TR-94-2241
SMU-DS-TR273

A HYPOTHESIS-TESTING APPROACH TO DISCRIMINANT ANALYSIS WITH MIXED CATEGORICAL AND CONTINUOUS VARIABLES WHEN DATA ARE MISSING

J.W. Miller
W.A. Woodward
H.L. Gray
G.D. McCartor

Southern Methodist University
Department of Statistical Science
Dallas, TX 75275

July 1994

Scientific Report No. 1

PHILLIPS LABORATORY
Directorate of Geophysics
AIR FORCE MATERIEL COMMAND
HANSCOM AIR FORCE BASE, MA 01731-3010

SPONSORED BY
Advanced Research Projects Agency (DoD)
Nuclear Monitoring Research Office
ARPA ORDER NO A128

MONITORED BY
Phillips Laboratory
CONTRACT NO. F19628-93-C-0199

This technical report has been reviewed and is approved for publication.

JAMES F. LEWKOWICZ
Contract Manager
Earth Sciences Division

JAMES F. LEWKOWICZ, Director
Earth Sciences Division

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify PL/TSI, 29 Randolph Road, Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No: 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1994	Statistical Report #1

4. TITLE AND SUBTITLE
A Hypothesis-Testing Approach to Discriminant Analysis with Mixed Categorical and Continuous Variables When Data Are Missing

5. FUNDING NUMBERS
F19628-93-C-0199
PE 62301E
PR NM93
TAGM WU AP

6. AUTHOR(S)
J.W. Miller
G.D. McCartor
W.A. Woodward
H.L. Gray

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Southern Methodist University
Department of Statistical Science
Dallas, TX 75275

8. PERFORMING ORGANIZATION REPORT NUMBER
SMU-DS-TR273

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Phillips Laboratory
29 Randolph Road
Hanscom AFB, Massachusetts 01731-3010

Contract Manager: James Lewkowicz, PL/GPEH

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
PL-TR-94-2241

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

In this report we consider the problem of discriminant analysis with discrete (categorical) and continuous variables with data missing at random. We use a hypothesis-testing approach based on the generalized likelihood ratio as proposed by Baek, et al (1994). We use bootstrapping to determine critical values in order to control the Type I error rate. We present three algorithms for dealing with this case, each assuming a different model for the data: (1) The INDICATOR algorithm replaces categorical variables with indicator variables, and treats these as if they were continuous; (2) the FULL algorithm assumes a multinomial distribution for the discrete part, and a multivariate normal distribution (with mean and covariances depending on the discrete part) as the conditional distribution of the continuous part given the discrete part; and (3) the COMMON algorithm assumes a multinomial distribution for the discrete part, and a multivariate normal distribution (with only the means depending on the discrete part) as the conditional distribution of the continuous part given the discrete part. (That is, a common covariance matrix is assumed across all multinomial cells.) The performance of these algorithms is compared through a simulation study. While the INDICATOR algorithm seems to have highest power, it also tends to display a higher Type I error rate than desired. The FULL and the COMMON algorithms have very similar power, but the COMMON algorithm appears to control the Type I error rate most effectively, and is least susceptible to problems occurring when some multinomial cells are sparsely represented.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Discriminant analysis, outlier detection, likelihood ratio tests, missing data.	42
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassicied	Unclassified	SAR

CONTENTS

1. Introduction

In Baek *et al* (1994) (subsequently abbreviated BGWMF) techniques are given for a hypothesis-testing approach to discriminant analysis in which one wishes to control one of the probabilities of misclassification. Methods are presented for continuous variables only, as well as for a mixture of continuous and categorical variables. Essentially, the hypothesis-testing approach based on the ratio of maximized likelihood functions proposed by Krzanowski (1980, 1982) is employed and the test statistic is bootstrapped in order to estimate critical values for the allocation rule in such a way that the error rate is controlled. In Miller *et al* (1993) (subsequently abbreviated (MGW)), a similar hypothesis-testing approach is used for discriminant analysis and outlier detection in the presence of missing data. The EM algorithm (Dempster *et al* (1977)) is employed to obtain maximum likelihood estimates of model parameters and compute the maximized likelihoods based on the available data. That paper, however, only considers the case in which all variables are continuous and, in fact, normally distributed.

In this report, we wish to consider the remaining case in which we have a mixture of continuous and categorical variables used as discriminants, and also missing data, potentially in both the training sets and in the new observation to be classified. Once again, we use a hypothesis-testing approach to classification and bootstrap the test statistic in order to control the probability of a particular type of misclassification. We present three algorithms for handling this situation:

(1) The INDICATOR algorithm - This algorithm begins by converting each categorical variable with j categories into j - 1 indicator variables. This results in a larger number of variables (unless all categorical variables are already binary, in which case the data set is unchanged). These indicator variables can be analyzed using techniques for quantitative data. In this algorithm we make the (obviously incorrect) assumption that all variates are continuous and, in fact, normally

distributed. We then perform discriminant analysis using the transformed data and the techniques of MGW.

(2) The FULL algorithm - Next, we model the joint distribution of each observation in the following manner: Suppose each observation consists of p categorical variables and q continuous variables. The categorical variables define r cells of a contingency table in which the observation could fall, where r is the product of the number of categories possible within each categorical variable. We assume that the observation will fall into cell i (i = 1, ..., r) with probability p_i, and that the conditional distribution of the continuous part given that the discrete part places the observation into cell i is multivariate normal with mean μ_i and Σ_i. We then employ the EM algorithm to obtain maximum likelihood estimates of parameters in this model and compute maximized likelihoods of the available data, and bootstrap the ratio of maximized likelihoods, as was done in BGWMF.

(3) The COMMON algorithm - This algorithm is essentially the same as the FULL algorithm, except that we assume a common covariance matrix across all multinomial cells. That is, the conditional distribution of the continuous part given that the discrete part places the observation into cell i is assumed multivariate normal with mean μ_i and Σ, with Σ no longer depending on i. This reduces considerably the number of parameters that need to be estimated and makes possible calculation of the likelihood ratio statistic when some cells may be sparsely represented, or not represented at all.

Simulation studies are conducted to compare and contrast the performance of each of these procedures with regard to their ability to accurately control the Type I error rate, and with regard to their power.

2. Notation and Overview of the Generalized Likelihood Ratio Test Procedure

Suppose we wish to classify a (p+q)-dimensional random vector **V** into one of two populations π_1 or π_2. Suppose further that **V** can be partitioned as **V** = (**X**, **Y**), where **X** = $(X_1, X_2, ..., X_p)$ is a p-dimensional vector of categorical variables and **Y** = $(Y_1, Y_2, ..., Y_q)$ is a q-dimensional vector of continuous variables. Suppose that for i = 1, ..., p, the variable X_i takes on one of the r_i possible values 1, 2, ..., r_i. Then the vector **X** takes on one of $r = \prod_{i=1}^{p} r_i$ possible values. We let Ψ denote the set of all possible values of the vector **X**. Finally, suppose that training samples $\{\mathbf{V}_i^{(1)}\}$, i = 1, ..., N_1 from π_1 and $\{\mathbf{V}_i^{(2)}\}$, i = 1, ..., N_2 from π_2, each having the same structure as **V**, are available, and that data may be missing at random from any part of **V** or from the training samples.

The generalized likelihood ratio test (GLRT) procedure for classifying **V** into π_1 or π_2 is based on a hypothesis testing approach. That is, the classification of **V** is done by testing

$$\begin{aligned} H_0&: \mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)} \in \pi_1; \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)} \in \pi_2 \\ &\qquad\qquad \text{versus} \qquad\qquad (1) \\ H_1&: \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)} \in \pi_1; \mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)} \in \pi_2. \end{aligned}$$

The two misclassification probabilities that we will be interested in are P(2|1) and P(1|2), where P(i|j) denotes the probability of classifying **V** into π_i when in fact $\mathbf{V} \in \pi_j$. We will refer to α = P(2|1) as the significance level for the test and P(2|2) as the power.

Let m denote the number of elements in **V** that are missing and let $\mathbf{V}_{(2)} = (\mathbf{X}_{(2)}, \mathbf{Y}_{(2)})$denote the (p - m)-variate vector of available data in **V**. Similarly, let $m_i^{(j)}$ denote the number of elements missing from $\mathbf{V}_i^{(j)}$ and let $\mathbf{V}_{i(2)}^{(j)}$ denote the $(p - m_i^{(j)})$-variate vector of available data in $\mathbf{V}_i^{(j)}$ (j = 1, 2; i = 1, 2, ..., N_j). We assume that π_1 has joint density function $f(\mathbf{V}|\theta^{(1)})$ and that π_2 has joint density function $f(\mathbf{V}|\theta^{(2)})$, where f is some

parametric density function with parameters $\theta^{(1)}$ and $\theta^{(2)}$ for populations π_1 and π_2, respectively. Then, under H_0, the likelihood of $\mathbf{V}$ and the training samples is given by

$$L_{01}(\theta^{(1)}|\ \mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) L_{02}(\theta^{(2)}|\ \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}), \tag{2}$$

where

$$L_{01}(\theta^{(1)}|\ \mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) = f_2(\mathbf{V}|\theta^{(1)}) \prod_{i=1}^{N_1} f_{1i}(\mathbf{V}_i^{(1)}|\theta^{(1)}), \tag{3}$$

$$L_{02}(\theta^{(2)}|\ \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}) = \prod_{i=1}^{N_2} f_{2i}(\mathbf{V}_i^{(2)}|\theta^{(2)}),$$

$f_2(\mathbf{V}|\theta^{(1)})$ is the marginal density function for $\mathbf{V}_{(2)}$ evaluated at $\mathbf{V}_{(2)}$ with parameters $\theta^{(1)}$, and $f_{ji}(\mathbf{V}_i^{(j)}|\theta^{(j)})$ is the marginal density function for $\mathbf{V}_{i(2)}^{(j)}$ evaluated at $\mathbf{V}_{i(2)}^{(j)}$ with parameters $\theta^{(j)}$. Under H_1, the likelihood of $\mathbf{V}$ and the training samples is given by

$$L_{11}(\theta^{(1)}|\ \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) L_{12}(\theta^{(2)}|\ \mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}), \tag{4}$$

where

$$L_{11}(\theta^{(1)}|\ \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) = \prod_{i=1}^{N_1} f_{1i}(\mathbf{V}_i^{(1)}|\theta^{(1)}),$$

$$L_{02}(\theta^{(2)}|\ \mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}) = f_2(\mathbf{V}|\theta^{(2)}) \prod_{i=1}^{N_2} f_{2i}(\mathbf{V}_i^{(2)}|\theta^{(2)}),$$

and $f_2(\mathbf{V}|\theta^{(2)})$ is the marginal density function for $\mathbf{V}_{(2)}$ evaluated at $\mathbf{V}_{(2)}$ with parameters $\theta^{(2)}$. We emphasize that these are the likelihood functions for the <u>available data</u> rather than the likelihood functions for the <u>complete data</u> since f_2 and f_{ji} ($j = 1, 2$; $i = 1, 2, \ldots, N_j$) are marginal densities for the available part of each observation, rather than the likelihood functions for the complete data.

The GLRT procedure is based on the ratio

$$LR = \frac{\sup_{(\theta^{(1)}, \theta^{(2)})} L_{01}(\theta^{(1)} | \mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) L_{02}(\theta^{(2)} | \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)})}{\sup_{(\theta^{(1)}, \theta^{(2)})} L_{11}(\theta^{(1)} | \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) L_{12}(\theta^{(2)} | \mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)})} \tag{5}$$

$$= \frac{L_{01}(\hat{\theta}_0^{(1)} | \mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) L_{02}(\hat{\theta}_0^{(2)} | \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)})}{L_{11}(\hat{\theta}_1^{(1)} | \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}) L_{12}(\hat{\theta}_1^{(2)} | \mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)})},$$

where $\hat{\theta}_0^{(j)}$ and $\hat{\theta}_1^{(j)}$ are maximum likelihood estimates of $\theta^{(j)}$ (j = 1, 2) under the null and alternative hypotheses, respectively. That is, $\hat{\theta}_0^{(1)}$ is the MLE of $\theta^{(1)}$ based on the sample $\{\mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}\}$, $\hat{\theta}_0^{(2)}$ is the MLE of $\theta^{(2)}$ based on the sample $\{\mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}\}$, and $\hat{\theta}_1^{(1)}$ is the MLE of $\theta^{(1)}$ based on the sample $\{\mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}\}$, and $\hat{\theta}_1^{(2)}$ is the MLE of $\theta^{(2)}$ based on the sample $\{\mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}\}$.

Equivalently, the test procedure may be based on the statistic

$$\lambda = \log(LR) = \lambda_{01} + \lambda_{02} - \lambda_{11} - \lambda_{12}, \tag{6}$$

where

$$\lambda_{01} = \log f_2(\mathbf{V} | \hat{\theta}_0^{(1)}) + \sum_{i=1}^{N_1} \log f_{1i}(\mathbf{V}_i^{(1)} | \hat{\theta}_0^{(1)}), \tag{7}$$

$$\lambda_{02} = \sum_{i=1}^{N_2} \log f_{2i}(\mathbf{V}_i^{(2)} | \hat{\theta}_0^{(2)}),$$

$$\lambda_{11} = \sum_{i=1}^{N_1} \log f_{1i}(\mathbf{V}_i^{(1)} | \hat{\theta}_1^{(1)}), \text{ and}$$

$$\lambda_{12} = \log f_2(\mathbf{V} | \hat{\theta}_1^{(2)}) + \sum_{i=1}^{N_1} \log f_{2i}(\mathbf{V}_i^{(2)} | \hat{\theta}_1^{(2)}).$$

A key step in evaluating λ for a given sample is the computation of the maximum likelihood estimates and the corresponding maximized log-likelihood functions λ_{01}, λ_{02}, λ_{11}, and λ_{12} in equation (7). This is no trouble when the data are complete, as illustrated by (BGWMF). However, in the presence of missing data, the usual expressions for

maximum likelihood estimates are no longer valid. In this case, maximum likelihood estimates are obtained via the EM algorithm (Dempster *et al* (1977)). The EM algorithm is an iterative procedure for obtaining parameter estimates which maximize the likelihood function of the available data. It involves two key steps:

(E -step) - Using current estimates $\hat{\theta}^{(k)}$ (where k now denotes the current iteration step, rather than designating π_1 or π_2), estimate the values of the complete data sufficient statistics by computing their expectations given the available data.

(M-step) - Determine the values of the parameters which maximize the likelihood for the complete data based on the current estimates of the complete data sufficient statistics, thus yielding $\hat{\theta}^{(k+1)}$.

The EM algorithm iteratively performs E- and M-steps until the sequence $\{\hat{\theta}^{(k)}\}$ converges to an adequate approximation to the MLE. To evaluate the test statistic λ of equation (3), we must implement the EM algorithm four times. That is, $\hat{\theta}_0^{(1)}$ and λ_{01} are based on the sample $\{\mathbf{V}, \mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}\}$, $\hat{\theta}_0^{(2)}$ and λ_{02} are based on the sample $\{\mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}\}$, $\hat{\theta}_1^{(1)}$ and λ_{11} are based on the sample $\{\mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}\}$, and $\hat{\theta}_1^{(2)}$ and λ_{12} are based on the sample $\{\mathbf{V}, \mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}\}$.

The decision rule is described as follows: small values of λ provide evidence in favor of H_1, hence, $\mathbf{V}$ is classified into π_2 if $\lambda \leq \lambda_\alpha$, otherwise $\mathbf{v}$ is classified into π_1. The cut-off value λ_α is chosen so that $P(2|1) = \alpha$, the desired significance level for the test. Since the null distribution of λ is not known, the critical value is approximated by the parametric bootstrap (Efron 1979). For some large integer B, B bootstrap samples $\{\mathbf{V}^*, \mathbf{V}_1^{*(1)}, \mathbf{V}_2^{*(1)}, \ldots, \mathbf{V}_{N_1}^{*(1)}\}$ are simulated from a distribution with density $f(\mathbf{V}|\hat{\theta}^{(1)})$ and B bootstrap samples $\{\mathbf{V}_1^{*(2)}, \mathbf{V}_2^{*(2)}, \ldots, \mathbf{V}_{N_2}^{*(2)}\}$ are simulated from a distribution with density $f(\mathbf{V}|\hat{\theta}^{(2)})$, where $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ are MLEs obtained from the samples $\{\mathbf{V}_1^{(1)}, \mathbf{V}_2^{(1)}, \ldots, \mathbf{V}_{N_1}^{(1)}\}$, and $\{\mathbf{V}_1^{(2)}, \mathbf{V}_2^{(2)}, \ldots, \mathbf{V}_{N_2}^{(2)}\}$, respectively. (Notice that in this case, $\hat{\theta}^{(1)} = \hat{\theta}_1^{(1)}$ and $\hat{\theta}^{(2)} =$

$\hat{\theta}_0^{(2)}$.) When there are missing values, the simulated bootstrap samples should also have missing values in a configuration similar to that in the actual data. For each bootstrap sample, the test statistic λ is computed, thus generating a random sample $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_B^*\}$ of variates that have approximately the same distribution as λ under H_0. For an α-level test, the cut-off value λ_α^* is chosen as the α-th empirical quantile of $\{\lambda_1^*, \lambda_2^*, \ldots, \lambda_B^*\}$. Finally, **V** is classified into π_1 if $\lambda > \lambda_\alpha^*$; **V** is classified into π_2 if $\lambda \leq \lambda_\alpha^*$.

As was pointed out in (MGW), this test procedure is only an approximation to the true GLRT procedure since the critical value is obtained via bootstrapping and we may further relax our approximation to the true GLRT procedure by relaxing the number of iterations performed by the EM algorithm. That is, we may choose a stopping criterion for the EM algorithm that does not continue iteration until convergence has been obtained to a high degree of accuracy. Whatever the stopping criterion, bootstrapping the test statistic insures an approximate α-level test. As in (MGW), it would appear that very little power is lost by only performing a very few iterations of the EM algorithm, as opposed to carrying out iterations until MLEs are obtained with a high degree of accuracy. In Section 6, we often use only three iterations as standard practice in our simulation studies.

Our implementation of the GLRT procedure for discriminant analysis is summarized in Figure 1. Figures 2, 3, and 4 further describe the bootstrapping module, the computation of the test statistic λ, and the EM algorithm for obtaining MLEs. Each of the various algorithms discussed in this report share this common skeletal structure. The differences lie in the type of model being assumed for the data, the corresponding implementation of the EM algorithm for obtaining MLEs, and the precise formulas used to evaluate the maximized log-likelihood functions.

3. The INDICATOR Algorithm

In our first attempt to implement an algorithm for discriminant analysis for mixed categorical and continuous variables with missing data, we desired to use the methods presented in MGW with as little adaptation as possible. One way to do this would be to treat the categorical variables as if they were continuous and use the procedures of MGW without any alteration at all. This is perhaps not such a bad idea if categorical variables have a large number of categories, if these categories have a natural ordering, and if the distribution of this variable has a somewhat normal shape. In most cases, however, these conditions are not satisfied and the procedure would be totally inappropriate.

A modification to the above approach is to replace each categorical variable X_i ($i = 1, \ldots, p$) from $\mathbf{V}$ with the $r_i - 1$ indicator variables

$$W_{ij} = \mathbf{I}(X_i = j) = \begin{cases} 1 \text{ if } X_i = j \\ 0 \text{ otherwise} \end{cases} \quad (j = 1, 2, \ldots, r_i - 1). \tag{8}$$

Hence, the vector $\mathbf{X}$ of categorical variables gets replaced by a vector $\mathbf{W}$ of binary variables of length $\Sigma_{i=1}^{p} r_i - p$, producing the transformed vector $\tilde{\mathbf{V}} = (\mathbf{W}, \mathbf{Y})$. If X_i is missing in $\mathbf{X}$, then each W_{ij} ($j = 1, 2, \ldots, r_i - 1$) is missing in $\mathbf{W}$. We transform the training samples $\mathbf{V}_i^{(j)}$ ($j = 1, 2$; $i = 1, 2, \ldots, N_j$) in a similar manner producing $\tilde{\mathbf{V}}_i^{(j)}$ ($j = 1, 2$; $i = 1, 2, \ldots, N_j$).

Now, having transformed each observation, we classify $\mathbf{V}$ by classifying $\tilde{\mathbf{V}}$ according to the GLRT procedure as outlined in (MGW) for the continuous-variables-only case with missing data using the transformed data $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{V}}_i^{(j)}$ ($j = 1, 2$; $i = 1, 2, \ldots, N_j$). That is, we proceed as if $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{V}}_i^{(j)}$ ($j = 1, 2$; $i = 1, 2, \ldots, N_j$) were normally distributed, ignoring the fact that many of the components are binary. In simulation studies (see Section 6, below), we see that this method actually performs about as well as

methods based on a more plausible model for the categorical variables, and is much easier to implement.

4. The FULL Algorithm

Next, we derive the GLRT procedure using a more plausible model for the distribution of $\mathbf{V}$. In this case, we assume that the distribution of $\mathbf{X}$ follows a multinomial distribution in the sense that $\Pr[\mathbf{X} = \mathbf{x}] = p_{\mathbf{x}}$ for each $\mathbf{x} \in \Psi$, and the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ is multivariate normal with mean $\mu_{\mathbf{x}}$ and covariance matrix and $\Sigma_{\mathbf{x}}$. Hence, $\theta = \{p_{\mathbf{x}}, \mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}; \mathbf{x} \in \Psi\}$ and

$$f(\mathbf{v}|\theta) = p_{\mathbf{x}}\mathrm{MVN}(\mathbf{y}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}}), \tag{9}$$

where $\mathrm{MVN}(\mathbf{y}|\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$ denotes the value of the multivariate normal density function with parameters $\mu_{\mathbf{x}}$ and $\Sigma_{\mathbf{x}}$ evaluated at $\mathbf{y}$.

The first step in deriving the GLRT procedure is to develop the EM algorithm for obtaining MLEs of θ given a collection of observations $(\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_n)$ with missing values from such a population. The vectors $\mathbf{V}_i$ may be partitioned as $(\mathbf{X}_i, \mathbf{Y}_i)$, where $\mathbf{X}_i$ and $\mathbf{Y}_i$ are the vectors of categorical variables, and continuous variables respectively, and further partitioned as $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{Y}_{1i}, \mathbf{Y}_{2i})$, where $\mathbf{X}_{1i}$ and $\mathbf{Y}_{1i}$ correspond to missing observations, and $\mathbf{X}_{2i}$ and $\mathbf{Y}_{2i}$ correspond to available observations. (In this final partitioning, the dimensions of the various pieces may vary with i, and elements may be permuted differently for each i according to the pattern of missing values in each observation.) We note that the complete-data sufficient statistics for the parameters in this model are

$$\begin{aligned} &N_{\mathbf{x}} = \Sigma_{i=1}^{n}\mathbf{I}(\mathbf{X}_i = \mathbf{x}), \\ &\mathbf{S}_{\mathbf{x}} = \Sigma_{i=1}^{n}\mathbf{I}(\mathbf{X}_i = \mathbf{x})\mathbf{Y}_i, \text{ and} \quad (\mathbf{x} \in \Psi) \\ &\mathbf{SS}_{\mathbf{x}} = \Sigma_{i=1}^{n}\mathbf{I}(\mathbf{X}_i = \mathbf{x})\mathbf{Y}_i\mathbf{Y}_i^{\mathrm{T}}, \end{aligned} \tag{10}$$

and that the MLEs for the parameters in this model based on the complete data are

$$\begin{aligned}
&\hat{p}_x = N_x/n, \\
&\hat{\mu}_x = S_x/N_x, \text{ and} \qquad (x \in \Psi) \\
&\hat{\Sigma}_x = SS_x/N_x - \hat{\mu}_x \hat{\mu}_x^T.
\end{aligned} \tag{11}$$

The M-step in this setting simply amounts to evaluating each of the pieces of (11). The main computational burden lies in computing the conditional expectations of the complete data sufficient statistics given the available data under current parameter estimates in each iteration (the E-step).

In the E-step, we wish to compute (under the distribution defined by $\hat{\theta}^{(k)}$)

$$\begin{aligned}
&E[N_x \mid \{(X_{2i}, Y_{2i}), i = 1, n\}] = \Sigma_{i=1}^{n} E[I(X_i = x) \mid (X_{2i}, Y_{2i})], \\
&E[S_x \mid \{(X_{2i}, Y_{2i}), i = 1, n\}] = \Sigma_{i=1}^{n} E[I(X_i = x)Y_i \mid (X_{2i}, Y_{2i})], \text{ and} \qquad (x \in \Psi) \\
&E[SS_x \mid \{(X_{2i}, Y_{2i}), i = 1, n\}] = \Sigma_{i=1}^{n} E[I(X_i = x)Y_i Y_i^T \mid (X_{2i}, Y_{2i})].
\end{aligned} \tag{12}$$

This computation is facilitated by the following identities:

$$E[I(X = x)h(Y) \mid (X_2, Y_2)] = Pr[X = x \mid (X_2, Y_2)] \cdot E[h(Y) \mid X = x, Y_2] \tag{13}$$

$$Pr[X = x \mid (X_2, Y_2)] = \frac{I(X_2 = x_2)p_x MVN(y_2 \mid \mu_x, \Sigma_x)}{\sum_{\tilde{x} \in \Psi} I(\tilde{x}_2 = X_2)p_{\tilde{x}} MVN(y_2 \mid \mu_{\tilde{x}}, \Sigma_{\tilde{x}})} \tag{14}$$

$$E[Y_1 \mid X = x, Y_2] = \mu_x^{(1)} + \Sigma_x^{(12)}\{\Sigma_x^{(22)}\}^{-1}(Y_2 - \mu_x^{(2)}) \tag{15}$$

$$E[Y_2 \mid X = x, Y_2] = Y_2 \tag{16}$$

$$\begin{aligned}
&E[Y_1 Y_i^T \mid X = x, Y_2] = \\
&\Sigma_x^{(11)} - \Sigma_x^{(12)}\{\Sigma_x^{(22)}\}^{-1}\Sigma_x^{(21)} + E[Y_1 \mid X = x, Y_2]E[Y_1 \mid X = x, Y_2]^T
\end{aligned} \tag{17}$$

$$E[Y_1 Y_2^T \mid X = x, Y_2] = E[Y_1 \mid X = x, Y_2] \cdot Y_2^T \tag{18}$$

$$E[\mathbf{Y}_2\mathbf{Y}_2^T \mid \mathbf{X} = \mathbf{x}, \mathbf{Y}_2] = \mathbf{Y}_2\mathbf{Y}_2^T \tag{19}$$

Here, $\mu_{\mathbf{x}}^{(1)}$, $\mu_{\mathbf{x}}^{(2)}$, $\Sigma_{\mathbf{x}}^{(11)}$, $\Sigma_{\mathbf{x}}^{(12)}$, $\Sigma_{\mathbf{x}}^{(21)}$, and $\Sigma_{\mathbf{x}}^{(22)}$ are appropriate partitions of $\mu_{\mathbf{x}}$ and $\Sigma_{\mathbf{x}}$ corresponding to the missing and available parts of $\mathbf{Y}$. Equations (15) - (19) are used to estimate the missing parts of $E[\mathbf{Y} \mid \mathbf{X} = \mathbf{x}, \mathbf{Y}_2]$ and $E[\mathbf{Y}\mathbf{Y}^T \mid \mathbf{X} = \mathbf{x}, \mathbf{Y}_2]$. Computation of the expectations in (12) is then carried out as follows: For each observation in the data set, compute $E[\mathbf{I}(\mathbf{X}_i = \mathbf{x}) \mid (\mathbf{X}_{2i}, \mathbf{Y}_{2i})]$, $E[\mathbf{I}(\mathbf{X}_i = \mathbf{x})\mathbf{Y}_i \mid (\mathbf{X}_{2i}, \mathbf{Y}_{2i})]$, and $E[\mathbf{I}(\mathbf{X}_i = \mathbf{x})\mathbf{Y}_i\mathbf{Y}_i^T \mid (\mathbf{X}_{2i}, \mathbf{Y}_{2i})]$ for each $\mathbf{x} \in \Psi$ via equations (13) - (19). Accumulate these over all observations to obtain (12).

In the case of continuous variables only, (MGW) used estimates of parameters based on substituting means for missing observations as initial estimates in the iterative process. This becomes more complicated in the presence of categorical variables. To simplify initialization, we use "blind initialization": we initialize each $p_{\mathbf{x}}$ with 1/r, each $\mu_{\mathbf{x}}$ with $\mathbf{0}$, and each $\Sigma_{\mathbf{x}}$ with $\mathbf{I}$. Experience so far indicates that the first iteration of the EM algorithm substantially alters the parameter estimates to something comparable to "mean substitution," if that means anything in this context. In any case, this initialization procedure has worked adequately so far in simulation studies.

Having evaluated the MLEs using the EM algorithm, we need a method for evaluating the maximized log-likelihood functions in equation (7). The likelihood function for the available data is the product of the likelihoods of the available parts of each observation. The likelihood of the available part of a single observation is the marginal density for (X_2, Y_2). This may be obtained from the density for (X, Y) by integrating out X_1 and Y_1. This gives

$$f_{X_2,Y_2}(\mathbf{x}_2,\mathbf{y}_2) = \sum_{\tilde{\mathbf{x}} \in \Psi} \mathbf{I}(\mathbf{x}_2 = \tilde{\mathbf{x}}_2)\, p_{\tilde{\mathbf{x}}}\, \mathrm{MVN}(\mathbf{y}_2 \mid \mu_{\tilde{\mathbf{x}}}, \Sigma_{\tilde{\mathbf{x}}}). \tag{20}$$

The maximized log-likelihood of the available data in a sample is obtained by accumulating the values of the log of (20) over all observations in the sample. Thus, we may evaluate each of the pieces λ_{01}, λ_{02}, λ_{11}, and λ_{12} in equation (7), from which we may evaluate the test statistic $\lambda = \log(LR)$ given in (6).

As will be seen in the simulation results of Section 6, the FULL Algorithm has one major flaw that must be addressed. That is, it can only be used in cases in which there is adequate representation in each cell to obtain a full-rank estimate of $\Sigma_{\mathbf{x}}$ for each $\mathbf{x} \in \Psi$ in both populations. If r is large, *i.e.*, if there are a large number of categorical variables, or a large number of categories within some categorical variables, or both, then the training samples may need to be extremely large so that all parameters can be estimated accurately. In practice, such large samples may not be available, and it becomes necessary to impose further constraints on the parameters of the model so that the number of parameters required is reduced. This leads us into our discussion of the next algorithm.

5. The COMMON Algorithm

This last algorithm is very similar to the FULL Algorithm except that in our model for the data, we assume that the conditional covariance matrix for the continuous part given the discrete part is common for all $\mathbf{x} \in \Psi$. That is, the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ is multivariate normal with mean $\mu_{\mathbf{x}}$ and covariance matrix and Σ not depending on $\mathbf{x}$. Hence, $\theta = \{p_{\mathbf{x}}, \mu_{\mathbf{x}}, \Sigma; \mathbf{x} \in \Psi\}$. This reduces the number of parameters that need to be estimated considerably, and makes parameter estimation possible when some cells are sparse, or not represented at all. We allow the possibility of different parameters for each of the two populations, but within each population, Σ is common across all multinomial cells. This model gives precisely the general location model of Olkin and Tate (1961). The EM algorithm for this model is developed by Little and Schluchter (1985), and they point out that this can be used to implement the GLRT

procedure proposed by Krzanowski (1982). What follows is precisely this procedure, with the added feature that we bootstrap the distribution of the test statistic in order to choose critical values to control the P(2|1) error rate. Although Little and Schluchter (1985) describe the EM algorithm for this model in considerable detail, we present a description of the algorithm here that is consistent with the notation of Section 4.

First, we observe that the complete-data sufficient statistics for the parameters in this model are

$$\begin{aligned} &N_{\mathbf{x}} = \Sigma_{i=1}^{n} \mathbf{I}(\mathbf{X}_i = \mathbf{x}), \\ &\mathbf{S}_{\mathbf{x}} = \Sigma_{i=1}^{n} \mathbf{I}(\mathbf{X}_i = \mathbf{x}) Y_i, \text{ and} \quad (\mathbf{x} \in \Psi) \\ &\mathbf{SS} = \Sigma_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^T, \end{aligned} \tag{21}$$

and that the MLEs for the parameters in this model based on the complete data are

$$\begin{aligned} &\hat{p}_{\mathbf{x}} = N_{\mathbf{x}}/n, \\ &\hat{\mu}_{\mathbf{x}} = \mathbf{S}_{\mathbf{x}}/N_{\mathbf{x}}, \text{ and} \quad (\mathbf{x} \in \Psi), \\ &\hat{\Sigma} = \Sigma_{\mathbf{x} \in \Psi} \hat{p}_{\mathbf{x}} \hat{\Sigma}_{\mathbf{x}}, \end{aligned} \tag{22}$$

where $\hat{\Sigma}_{\mathbf{x}}$ is given by equations (10) and (11). In other words, the MLE of Σ in this case is precisely a weighted average of MLEs of $\Sigma_{\mathbf{x}}$ for each $\mathbf{x}$ based on the FULL model with weights $\hat{p}_{\mathbf{x}}$. Hence, we may perform the M-step in this algorithm with exactly the same formulas as the M-step in the FULL Algorithm, except that after each $\hat{\Sigma}_{\mathbf{x}}$ is computed, we average these according to (22) to obtain the updated estimate of the common Σ. The E-step for this algorithm is also identical to the E-step in the FULL Algorithm, except that throughout formulas (13) - (19), each $\hat{\Sigma}_{\mathbf{x}}$ is replaced by the common $\hat{\Sigma}$. The evaluation of the maximized log-likelihood functions in (7) is also performed using (20), as in the FULL algorithm, with again, the only difference being that each $\hat{\Sigma}_{\mathbf{x}}$ is replaced by the common $\hat{\Sigma}$.

6. Simulation Results

We have performed simulations of each of the three algorithms (INDICATOR, FULL, and COMMON) based on several different parameter configurations in order to determine how well the algorithm controls the Type I misclassification probability as desired, and to assess the power P(2|2) of each algorithm. We also keep track of how many times the algorithm fails to classify the observation at all. These failures occur when for some reason the simulated data fails to yield full-rank estimates of all required covariance matrix parameters. This results in an undefined test statistic λ. This happens most frequently in the FULL algorithm, and is caused by a very few number of observations falling into one or more of the multinomial cells. It happens occasionally in the INDICATOR algorithm when at least one possible value of a categorical variable is not represented. Failures may occur when the test statistic is undefined for the sample which we are trying to classify, and also when the statistic is undefined for attempted bootstrap samples. We see in our simulations that the COMMON algorithm is least susceptible to these types of failures.

Our first simulation involved the same parameter configurations used in BGWMF (Case 3: Mixture of Categorical and Continuous Variables). That is, we consider the case in which the categorical part is a single Bernoulli variable and the continuous part is a single normal random variable independent of the categorical variable. For population π_1, the Bernoulli parameter is $p_1 = 0.1$. The mean and variance of the continuous variable are $\mu_1 = 0$ and $\sigma_1^2 = 0.5$, respectively. For population π_2, we use $p_2 = 0.9$, 0.7, and 0.5, $\sigma_2^2 = 1.0$, and $\mu_2 = 0.5 + \Delta\sigma_2^2$ where Δ takes on values 0, 1, 2, and 3. The observed significance level $\hat{P}(2|1)$ is the proportion of times out of 500 simulated trials in which the variable **V** is classified into π_2 when, in fact, it was simulated from π_1. The estimated power $\hat{P}(2|2)$ is the proportion of times out of 500 simulated trials in which the variable **V** is classified into π_2 when, in fact, it was simulated from π_2. In order to achieve an approximate significance level of $\alpha = 0.05$, the variable **V** was classified into

π_2 if the test statistic λ is less than or equal to λ^*_α, the 0.05-th empirical quantile of $\{\lambda^*_1, \lambda^*_2, \ldots, \lambda^*_B\}$.

For $N_1 = N_2 = 50$ and $B = 99$, the power estimates are plotted in Figure 5, based on simulations with no missing data. We see that the FULL and COMMON algorithms agree very well with the power curves plotted in BGWMF (Figure 2). In fact, with no missing data, the COMMON algorithm is essentially equivalent to the method of BGWMF, so these simulation results should agree very well, as they do. The INDICATOR algorithm does not agree well with the FULL and COMMON algorithms. For this reason, the points corresponding to the INDICATOR algorithm are not connected with lines, since this would clutter the plot. It would seem that the INDICATOR algorithm has higher power in general than the other two. This is surprising since this algorithm does not model well the true distribution of the binary variable. A closer examination of the simulation results shows that this is, in fact, misleading, since the INDICATOR tends to yield a significance level nearly twice the desired 0.05 level. This can be seen in Figure 6, which shows the power estimate plotted versus the observed significance level. Each plot in Figure 6 corresponds to a specific value of Δ. We can also see that the COMMON algorithm most accurately achieves the desired $\alpha = 0.05$ significance level.

In Figures 7 and 8, we show corresponding plots based on data with missing values. In these simulations, each variable in each observation was deleted independently with probability 0.1, so that roughly 10% of the data is missing. We see an overall decrease in the power of all three algorithms compared to the full-data case, but this is to be expected since the test is based on less available data. Otherwise, the results of the missing-data case are comparable to the results of the full-data case.

We have tabulated the results of this simulation in Table 1. The ERROR column shows the percentage of times out of the 500 simulations that the algorithm failed to classify **V** due to singular parameter estimates. We see that the FULL algorithm is most

NO MISSING DATA						10 PERCENT MISSING DATA					
PROG	DELTA	P2	SIG	PWR	ERR	PROG	DELTA	P2	SIG	PWR	ERR
I	0	0.9	0.1006	0.8893	0.6	I	0	0.9	0.0872	0.8661	1.4
I	0	0.7	0.0946	0.6881	0.6	I	0	0.7	0.0884	0.6406	0.4
I	0	0.5	0.0946	0.5191	0.6	I	0	0.5	0.0823	0.4940	0.4
I	1	0.9	0.0926	0.8994	0.6	I	1	0.9	0.0791	0.8905	1.4
I	1	0.7	0.0905	0.8089	0.6	I	1	0.7	0.0843	0.7952	0.4
I	1	0.5	0.0885	0.7183	0.6	I	1	0.5	0.0763	0.6888	0.4
I	2	0.9	0.0885	0.9638	0.6	I	2	0.9	0.0791	0.9513	1.4
I	2	0.7	0.0885	0.9376	0.6	I	2	0.7	0.0805	0.9155	0.6
I	2	0.5	0.0885	0.9175	0.6	I	2	0.5	0.0803	0.8755	0.4
I	3	0.9	0.0845	0.9980	0.6	I	3	0.9	0.0730	0.9858	1.4
I	3	0.7	0.0825	0.9940	0.6	I	3	0.7	0.0763	0.9799	0.4
I	3	0.5	0.0825	0.9920	0.6	I	3	0.5	0.0763	0.9779	0.4
F	0	0.9	0.0681	0.6085	6.0	F	0	0.9	0.0561	0.5611	0.2
F	0	0.7	0.0638	0.4938	2.8	F	0	0.7	0.0481	0.4609	0.2
F	0	0.5	0.0779	0.3668	2.4	F	0	0.5	0.0481	0.3507	0.2
F	1	0.9	0.0581	0.7828	7.0	F	1	0.9	0.0601	0.7896	0.2
F	1	0.7	0.0723	0.7190	3.2	F	1	0.7	0.0501	0.6954	0.2
F	1	0.5	0.0717	0.6721	2.4	F	1	0.5	0.0561	0.6353	0.2
F	2	0.9	0.0746	0.9616	6.2	F	2	0.9	0.0520	0.9420	0.0
F	2	0.7	0.0825	0.9340	3.0	F	2	0.7	0.0500	0.8920	0.0
F	2	0.5	0.0799	0.9139	2.4	F	2	0.5	0.0540	0.8660	0.0
F	3	0.9	0.0423	0.9937	5.4	F	3	0.9	0.0582	0.9799	0.4
F	3	0.7	0.0887	0.9938	3.0	F	3	0.7	0.0501	0.9780	0.2
F	3	0.5	0.0799	0.9877	2.4	F	3	0.5	0.0541	0.9739	0.2
C	0	0.9	0.0480	0.5920	0.0	C	0	0.9	0.0440	0.5700	0.0
C	0	0.7	0.0480	0.4640	0.0	C	0	0.7	0.0480	0.4520	0.0
C	0	0.5	0.0460	0.3660	0.0	C	0	0.5	0.0520	0.3520	0.0
C	1	0.9	0.0520	0.8540	0.0	C	1	0.9	0.0360	0.8080	0.0
C	1	0.7	0.0500	0.7500	0.0	C	1	0.7	0.0440	0.7140	0.0
C	1	0.5	0.0480	0.6920	0.0	C	1	0.5	0.0440	0.6500	0.0
C	2	0.9	0.0460	0.9640	0.0	C	2	0.9	0.0440	0.9420	0.0
C	2	0.7	0.0480	0.9440	0.0	C	2	0.7	0.0440	0.8980	0.0
C	2	0.5	0.0380	0.9240	0.0	C	2	0.5	0.0440	0.8640	0.0
C	3	0.9	0.0500	1.0000	0.0	C	3	0.9	0.0440	0.9840	0.0
C	3	0.7	0.0500	0.9940	0.0	C	3	0.7	0.0460	0.9760	0.0
C	3	0.5	0.0400	0.9940	0.0	C	3	0.5	0.0440	0.9780	0.0

PROGRAM CODE: I - INDICATOR F - FULL C - COMMON

Table 1. Listing of results from first simulation.

susceptible to failures of this sort, failing as much as 6 to 7% of the time when $p_2 = 0.9$ and no data is missing. The INDICATOR algorithm failed somewhat less frequently, and the COMMON algorithm never failed in this study.

In our next simulation study, we consider the case in which we have two categorical variables, each with two categories, and two continuous variables. For population π_1, each possible combination of the categorical part ($\mathbf{X}$ = (1,1), (1,2), (2,1) and (2,2)) occurs with probability 1/4. The conditional distribution of the continuous part is MVN($\mathbf{0}$, Σ_1), where

$$\Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \tag{23}$$

within each multinomial cell (*i.e.*, conditional on each possible value of the discrete part). For population π_2, the conditional covariance matrix for the continuous part is Σ_2, where Σ_2 is given by

$$\Sigma_2 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}. \tag{24}$$

We use three different probability distributions for the discrete part, and four different configurations of mean vectors for the conditional distributions of the continuous part given each possible discrete part. In the plots and tables which follow, the three probability distributions are coded with the variable PCODE, which takes on values 1, 2, and 3. The four mean vector configurations are coded with the variable MCODE, which takes on values 1, 2, 3, and 4. The parameter configurations defined by these codes are shown in Table 2.

PCODE	Pr[**X** = (1, 1)]	Pr[**X** = (1, 2)]	Pr[**X** = (2, 1)]	Pr[**X** = (2, 2)]
1	0.25	0.25	0.25	0.25
2	0.50	0.20	0.20	0.10
3	0.80	0.10	0.10	0.00
MCODE	E[**Y**|**X** = (1, 1)]	E[**Y**|**X** = (1, 2)]	E[**Y**|**X** = (2, 1)]	E[**Y**|**X** = (2, 2)]
1	(0,0)	(0,0)	(0,0)	(0,0)
2	(2,2)	(1,1)	(1,1)	(0,0)
3	(0,0)	(1,1)	(1,1)	(2,2)
4	(2,2)	(2,2)	(2,2)	(2,2)

Table 2. Definitions for parameter codes used in out second simulation study.

PCODE = 1 corresponds to a uniform distribution across all multinomial cells. PCODE = 2 and PCODE = 3 correspond to distributions increasingly favoring cell (1,1). MCODE = 1 corresponds to a mean configuration identical to that for population π_1. MCODE = 2 and MCODE = 3 correspond to changes in mean for certain cells, and MCODE = 4 corresponds to the sum of these two changes. For PCODE = 1 and MCODE = 1, population π_2 is identical to population π_1 except for the correlation between the two continuous variables.

As in our first study, we take $N_1 = N_2 = 50$, B = 99, $\alpha = 0.05$, and base our observed significance level and power estimates on 500 replications of the procedure in each case. Figure 9 shows the power estimates plotted versus the mean configuration when no data is missing. Figure 10 shows plots of the power estimate versus the observed significance level for each mean configuration. Figures 11 and 12 are corresponding plots for approximately 10% missing data, with data deleted at random in the same manner as our previous study. Table 3 shows a listing of these results, including the percentages of failures due to singular parameter estimates.

In Figure 9, we see the power increases in general as the separation between in means increases (*i.e.*, as MCODE changes from 1 to 4). MCODE = 2 and MCODE = 3

NO MISSING DATA						10 PERCENT MISSING DATA					
PROG	PCODE	MCODE	SIG	PWR	ERR	PROG	PCODE	MCODE	SIG	PWR	ERR
I	1	1	0.0400	0.2300	0.0	I	1	1	0.0340	0.1860	0.0
I	2	1	0.0580	0.2520	0.0	I	2	1	0.0380	0.2220	0.0
I	3	1	0.1091	0.5859	1.0	I	3	1	0.1084	0.5072	2.2
I	1	2	0.0640	0.4540	0.0	I	1	2	0.0500	0.3820	0.0
I	2	2	0.0560	0.5920	0.0	I	2	2	0.0520	0.5360	0.0
I	3	2	0.0889	0.8283	1.0	I	3	2	0.0757	0.8200	2.2
I	1	3	0.0460	0.4380	0.0	I	1	3	0.0340	0.3720	0.0
I	2	3	0.0500	0.3080	0.0	I	2	3	0.0480	0.2540	0.0
I	3	3	0.1172	0.5657	1.0	I	3	3	0.1207	0.4806	2.2
I	1	4	0.0720	0.8440	0.0	I	1	4	0.0660	0.7840	0.0
I	2	4	0.0700	0.8540	0.0	I	2	4	0.0740	0.7980	0.0
I	3	4	0.0788	0.9010	1.0	I	3	4	0.0838	0.8875	2.2
F	1	1	0.0480	0.1660	0.0	F	1	1	0.0460	0.1720	0.0
F	2	1	0.0460	0.1926	8.6	F	2	1	0.0425	0.1680	1.2
F	3	1	0.0507	0.2230	40.8	F	3	1	0.0602	0.2229	66.8
F	1	2	0.0541	0.3627	0.2	F	1	2	0.0600	0.3360	0.0
F	2	2	0.0576	0.5543	9.8	F	2	2	0.0468	0.4725	1.8
F	3	2	0.0766	0.8029	45.2	F	3	2	0.0637	0.6115	68.6
F	1	3	0.0441	0.4309	0.2	F	1	3	0.0460	0.3040	0.0
F	2	3	0.0500	0.3420	9.2	F	2	3	0.0489	0.2688	1.8
F	3	3	0.0547	0.3029	45.2	F	3	3	0.0600	0.1667	70.0
F	1	4	0.0700	0.7880	0.0	F	1	4	0.0560	0.7520	0.0
F	2	4	0.0625	0.8060	7.2	F	2	4	0.0848	0.7475	1.0
F	3	4	0.0871	0.9024	42.6	F	3	4	0.0325	0.6558	69.2
C	1	1	0.0480	0.2400	0.0	C	1	1	0.0400	0.1860	0.0
C	2	1	0.0580	0.2660	0.0	C	2	1	0.0340	0.2100	0.0
C	3	1	0.0440	0.3360	0.0	C	3	1	0.0400	0.2800	0.0
C	1	2	0.0380	0.4680	0.0	C	1	2	0.0360	0.4020	0.0
C	2	2	0.0440	0.6080	0.0	C	2	2	0.0380	0.5620	0.0
C	3	2	0.0540	0.8180	0.0	C	3	2	0.0560	0.7660	0.0
C	1	3	0.0580	0.4840	0.0	C	1	3	0.0500	0.4140	0.0
C	2	3	0.0460	0.3380	0.0	C	2	3	0.0380	0.2840	0.0
C	3	3	0.0480	0.3400	0.0	C	3	3	0.0400	0.2880	0.0
C	1	4	0.0580	0.8380	0.0	C	1	4	0.0560	0.7980	0.0
C	2	4	0.0520	0.8720	0.0	C	2	4	0.0620	0.8100	0.0
C	3	4	0.0460	0.9180	0.0	C	3	4	0.0560	0.8740	0.0

PROGRAM CODE: I - INDICATOR F - FULL C - COMMON

Table 3. Listing of results from second simulation study.

actually correspond to the same degree of difference in means, so the power for these are not expected to be too different. In fact, the power estimates for MCODE = 2 and MCODE = 3 are very similar when PCODE = 1. However, they are not very similar when PCODE = 2 or 3. In this case, power is lower for MCODE = 3 than for MCODE = 2. This results since for MCODE = 3, the means differ in sparse cells, whereas for MCODE = 2, the means differ most in the most common cell (corresponding to $\mathbf{X}$ = (1,1)), making it more easy to differentiate between the two populations. We see similar patterns in Figure 11, and can also see a general decrease in power due to missing values.

These plots seem to indicate that the INDICATOR and COMMON algorithms have very similar power, these being generally better than the FULL algorithm. As in our first study, we notice in Figures 10 and 12 that the INDICATOR algorithm has a tendency to yield a higher significance level than desired, especially when PCODE = 3 (*i.e.*, when some cells are very sparse).

We see from Table 3 that the INDICATOR algorithm fails occasionally due to singular covariance matrix, especially when PCODE = 3. The FULL algorithm does much worse when cells are sparse. The FULL algorithm fails about two-thirds of the time when PCODE = 3 and data is missing! When some cells occur with very low probability, it is necessary to have very large samples so that each cell is represented enough to obtain a full-rank estimate of the covariance matrix within that cell. Samples of size 50, cells are not adequately represented about two-thirds of the time. Once again, we see that the COMMON algorithm is least susceptible to failures due to singular parameter estimates.

Readers may wonder why the algorithm doesn't fail *every time* for PCODE = 3 since cell (1,1) is *never* represented. If a cell probability is estimated to be zero, the covariance matrix estimate for that cell is never used in the computation of λ, and can, therefore, be disregarded. It is not cell (1,1) that is the problem here, rather it is cells (0,1) and (1,0). Readers may also find it strange that for PCODE = 2, there are fewer

failures in the FULL algorithm when data is missing than when all data is available. This may be explained intuitively as follows: When data is missing in the discrete part, there is some possibility that the observation falls into any of a number of cells. This observation contributes to the parameter estimates for all cells to which the observation might truly belong, resulting in fewer rank problems in sparse cells.

7. Concluding Remarks

In this report, we have extended the results of BGWMF and MGW to perform discriminant analysis with categorical and continuous variables when data is missing. We presented three algorithms for doing so. In simulation studies, we have observed that the INDICATOR algorithm has a tendency to yield a higher Type I error rate than desired. The FULL algorithm often fails due to singular parameter estimates when some value of the discrete part is sparsely represented. The COMMON algorithm seems to avoid these problems, and is, therefore, the preferred algorithm, especially when samples are small and the assumption of a common covariance matrix across all multinomial cells is reasonable. The code has now been transferred to MRC and Dr. Mark Fisk is applying these techniques to some existing seismic data.

References

Baek, J., Gray, H. L., Woodward, W. A., Miller. J. W., and Fisk, M. D. (1994), "A Bootstrap Generalized Likelihood Ratio Test in Discriminant Analysis," submittted to *Computational Statistics and Data Analysis*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood Estimation from Incomplete Data via the *EM* Algorithm" with discussion, *J. Royal Stat. Soc.*, ,1-38.

Efron, B., (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Little, R. J. A. & Schluchter, M. D. (1985) Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika* 72, 497-512.

Krzanowski, W. J. (1980) Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics* 36, 493-499.

Krzanowski, W. J. (1982) Mixtures of continuous and categorical variables in discriminant analysis: a hypothesis-testing approach. *Biometrics* 38, 991-1002.

Miller, J. W., Gray, H. L., and Woodward, W. A. (1993), "A Bootstrap Generalized Likelihood Ratio Test in Discriminant Analysis." Advanced Research Projects Agency Technical Report #3, Contract #F29601-91-K-DB25.

Olkin, I., & Tate, R. F. (1961) Multivariate correlation models with mixed discrete and continuous variables. *Ann. Math. Statist.* 32, 448-65

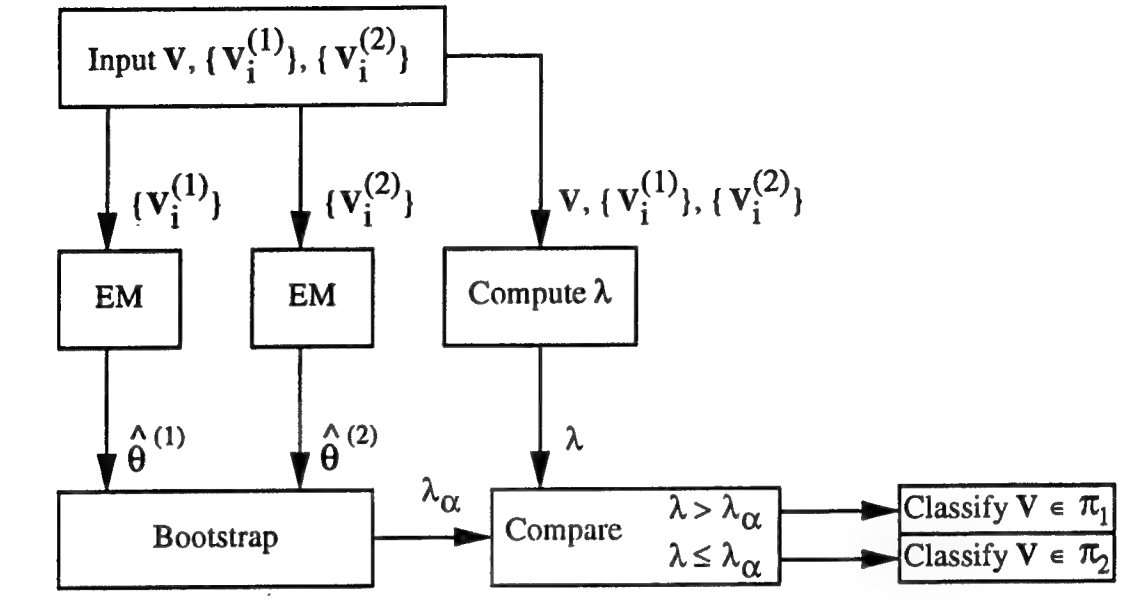


Figure 1. Flowchart for the GLRT procedure for discriminant analysis

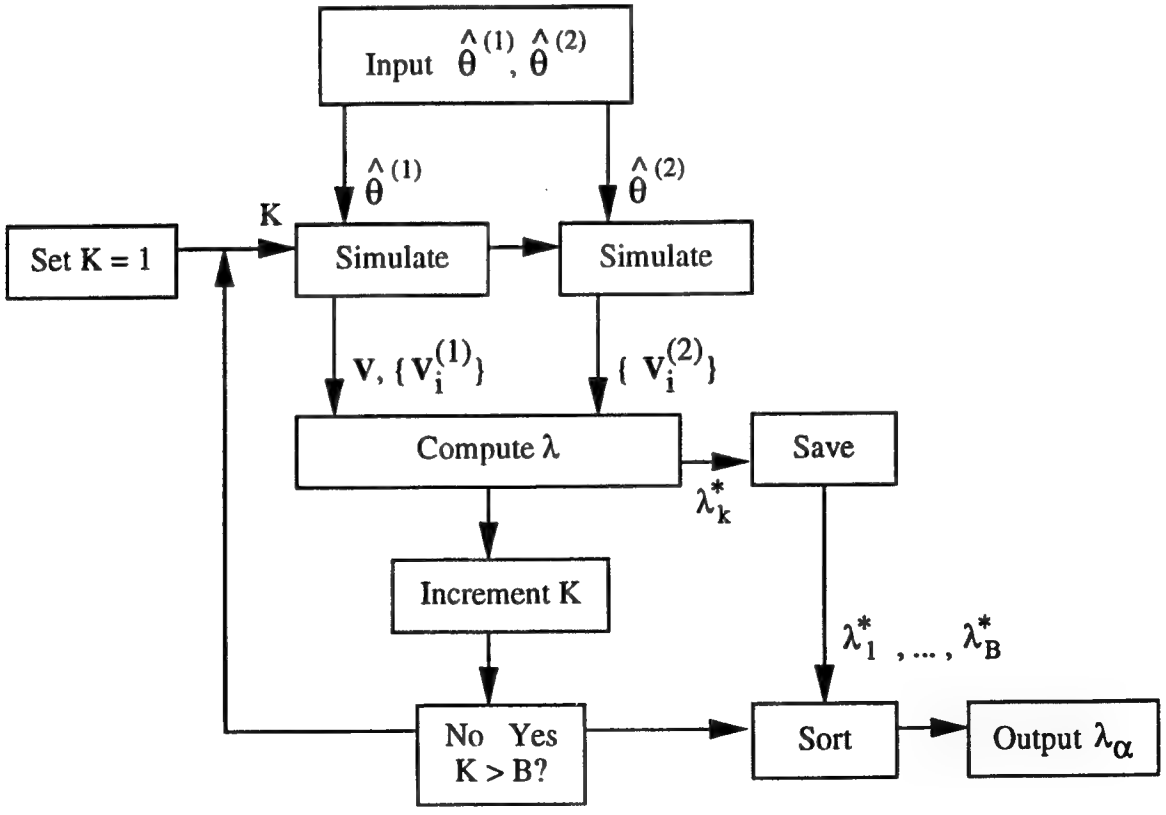


Figure 2. Flowchart for the bootstrapping portion of the GLRT procedure.

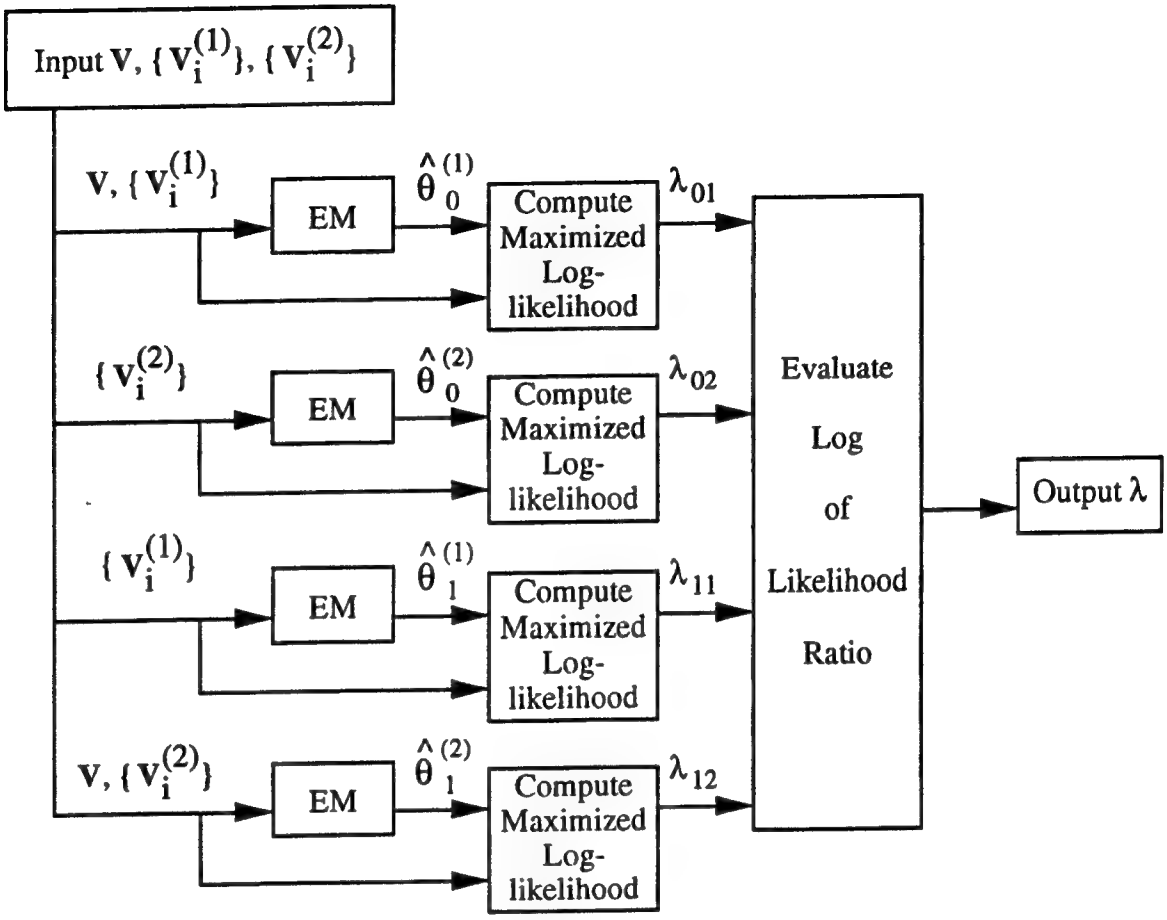


Figure 3. Flowchart for computing the test statistic λ in the GLRT procedure.

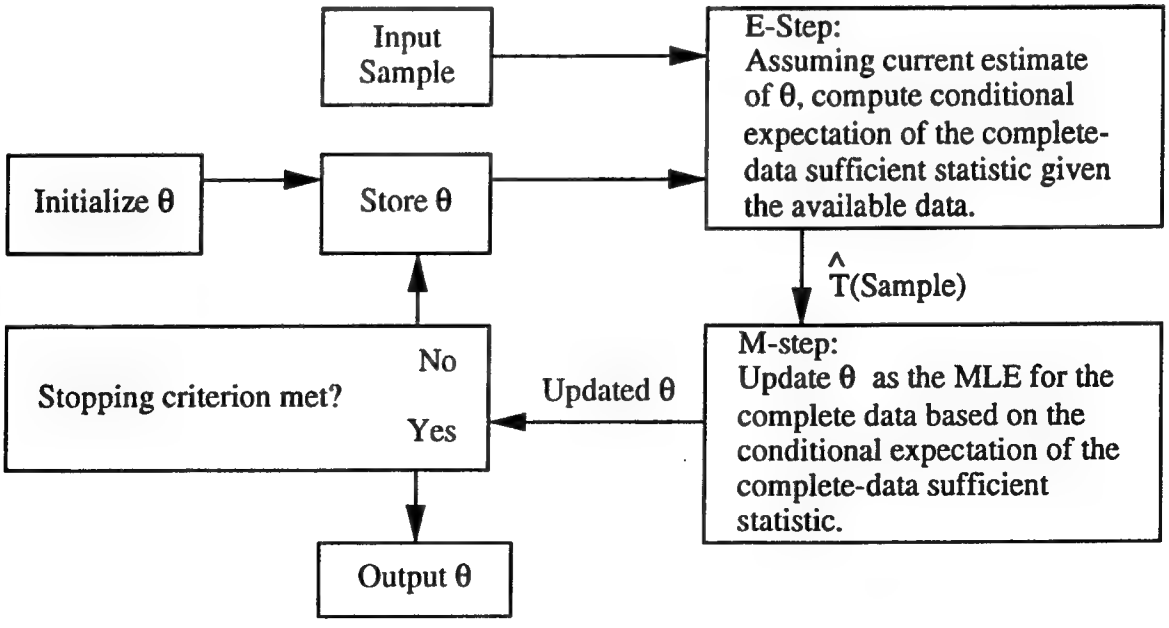


Figure 4. Flowchart for the EM algorithm.

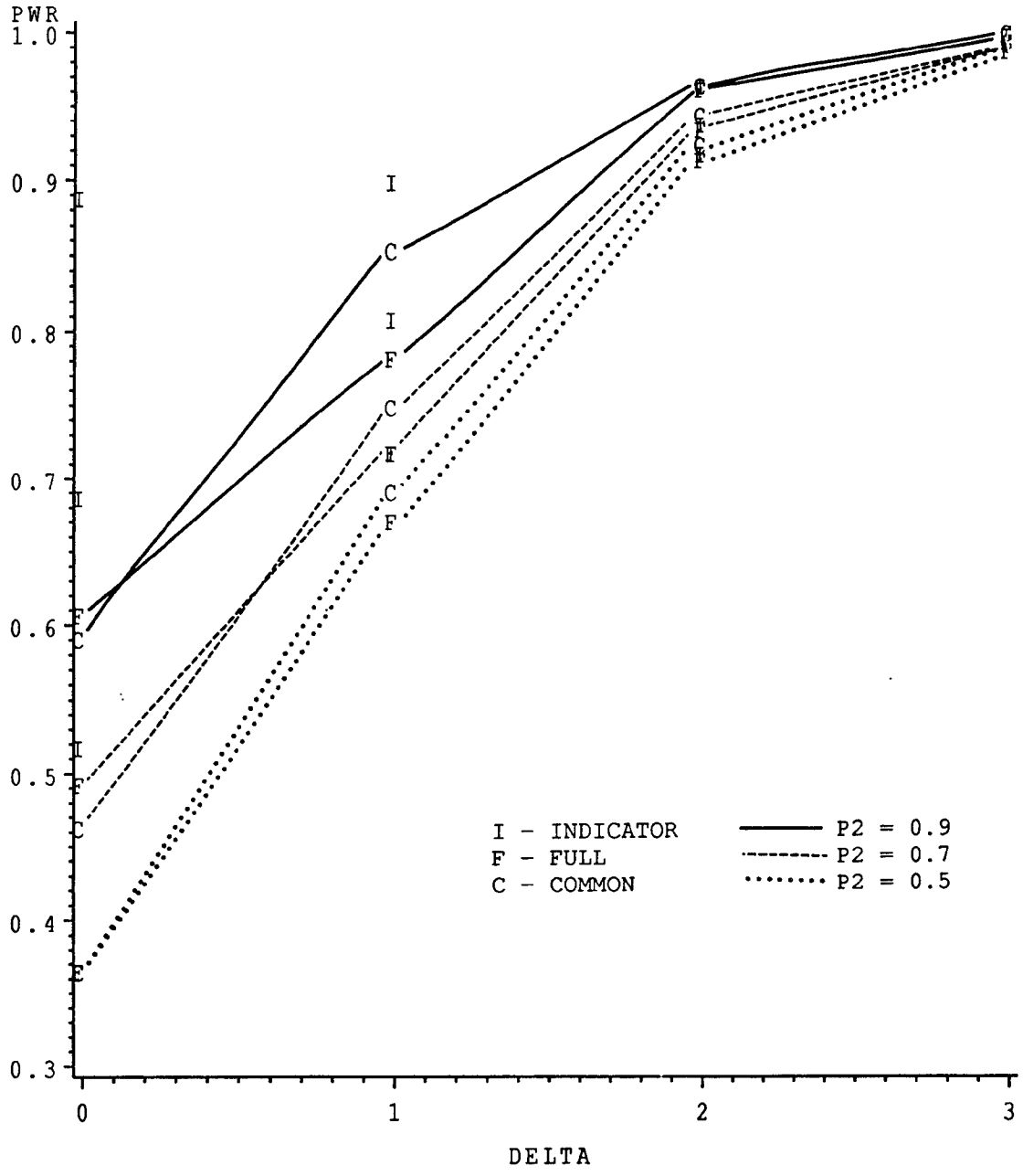
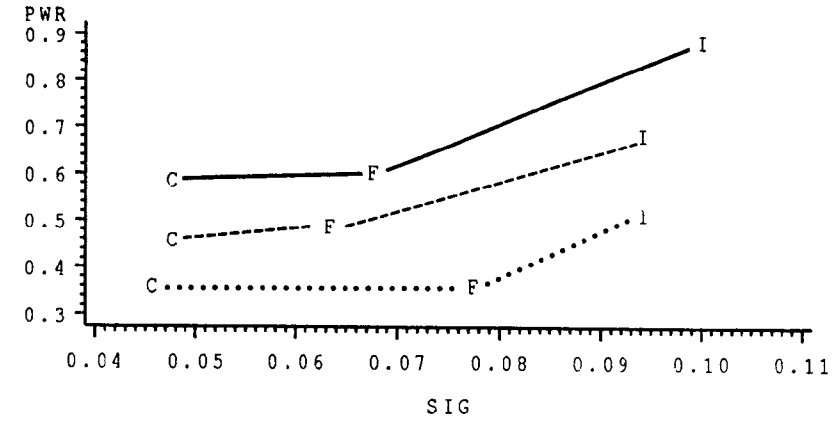
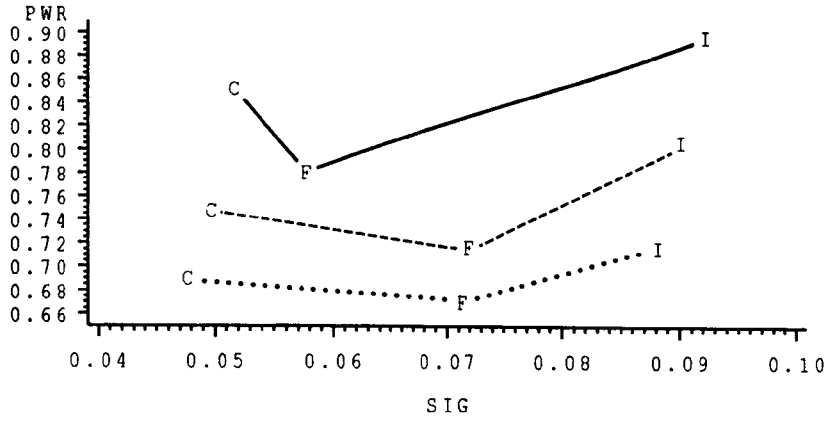


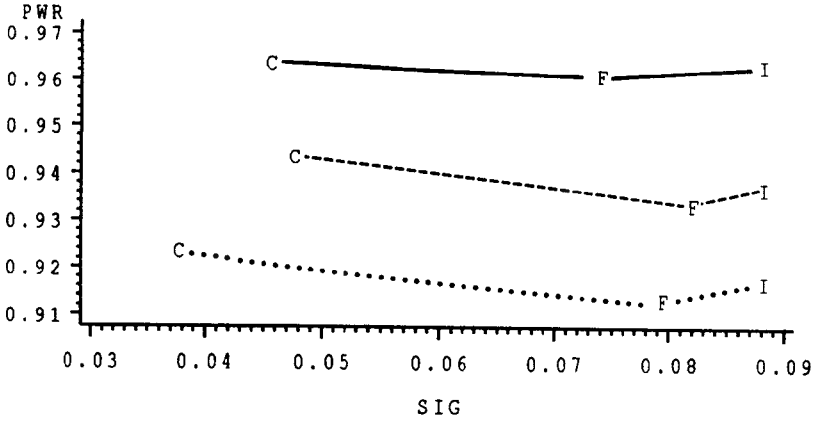
Figure 5. Power estimates when no data is missing for each of the three algorithms, with data modeled as a Bernoulli random variable and an independent normal random variable. Parameters for population π_1 are $p_1 = 0.1$, $\mu_1 = 0$, and $\sigma_1^2 = 0.5$. Power estimates are based on the following configurations for population π_2: $p_2 = 0.9$, 0.7, and 0.5, $\sigma_2^2 = 1.0$, and $\mu_2 = 0.5 + \Delta\sigma_2^2$, where Δ takes on values 0, 1, 2, and 3. For each value of Δ, the symbols I, F, and L are plotted at the corresponding power.



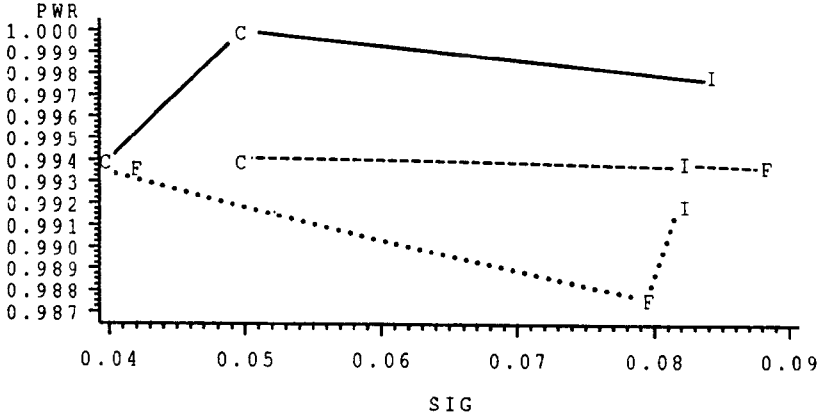
(a) $\Delta = 0$



(b) $\Delta = 1$



(c) $\Delta = 2$



(d) $\Delta = 3$

Figure 6. Plots of estimated power versus observed significance for each value of Δ when no data is missing. Symbol used is I for INDICATOR algorithm, F for FULL, and C for COMMON. The solid line corresponds to $p_2 = 0.9$, the dashed line to $p_2 = 0.7$, and the dotted line to $p_2 = 0.5$. The symbols I, F, and C are plotted at positions associated with the ordered pairs (significance level, power). The lines are not indicative of significance level or power but are only drawn to identify the ordered pairs associated with a particular value of p_2.

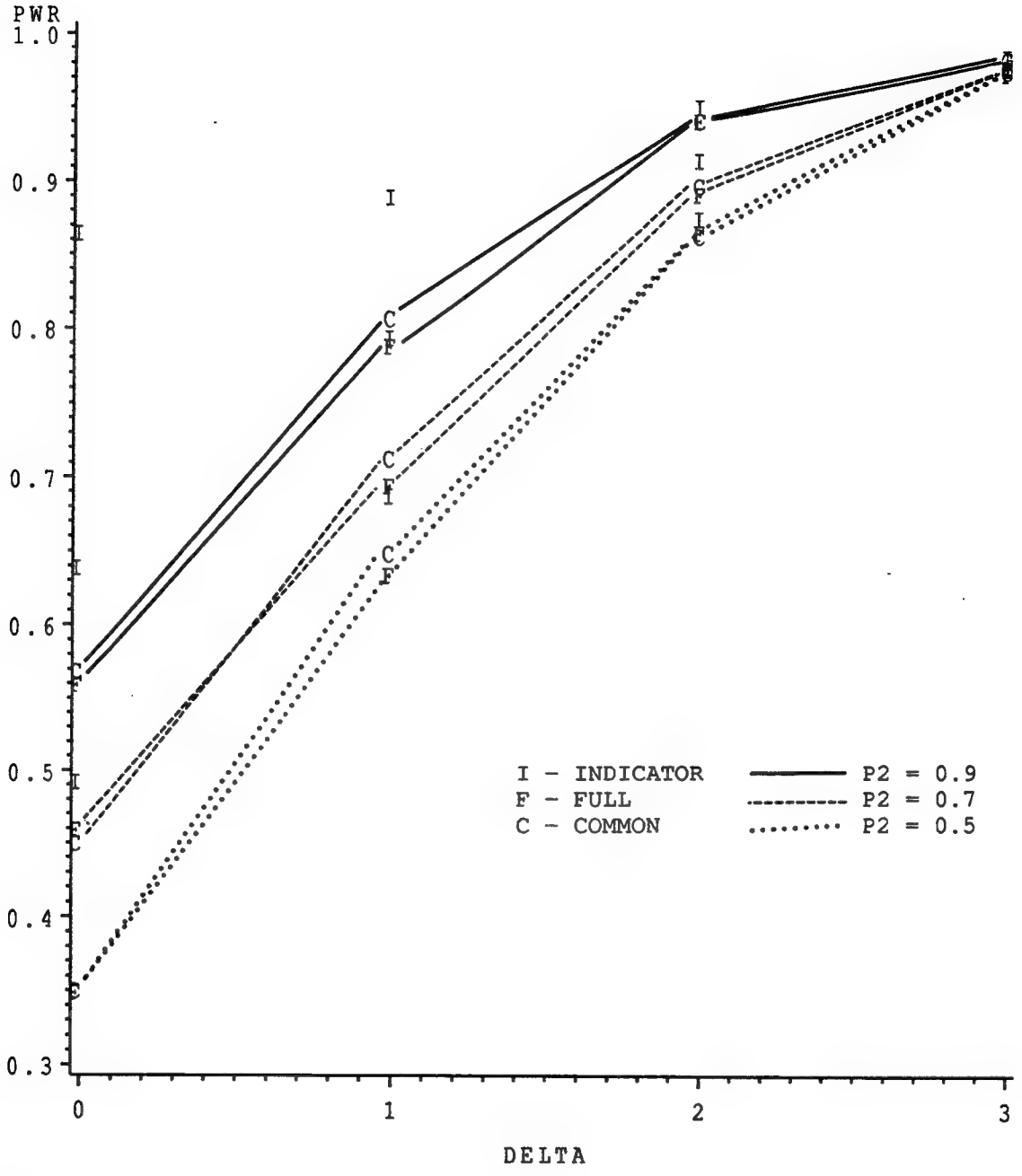


Figure 7. Power estimates when approximately 10% of the data is missing for each of the three algorithms, with data modeled as a Bernoulli random variable and an independent normal random variable. Parameters for population π_1 are $p_1 = 0.1$, $\mu_1 = 0$, and $\sigma_1^2 = 0.5$. Power estimates are based on the following configurations for population π_2: $p_2 = 0.9$, 0.7, and 0.5, $\sigma_2^2 = 1.0$, and $\mu_2 = 0.5 + \Delta\sigma_2^2$, where Δ takes on values 0, 1, 2, and 3.

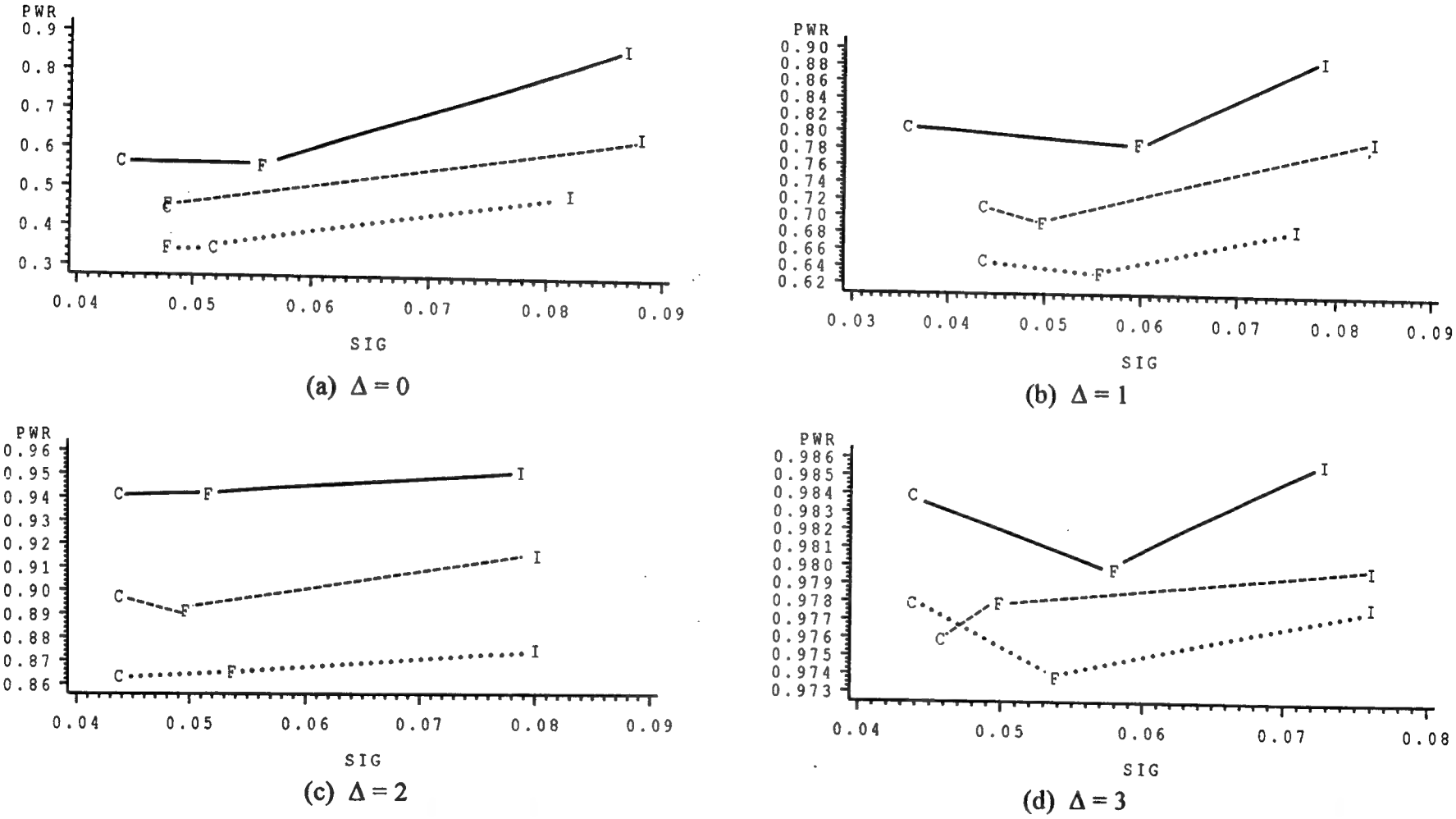


Figure 8. Plots of estimated power versus observed significance for each value of Δ when 10% of the data is missing. Symbol used is I for INDICATOR algorithm, F for FULL, and C for COMMON. The solid line corresponds to $p_2 = 0.9$, the dashed line to $p_2 = 0.7$, and the dotted line to $p_2 = 0.5$.

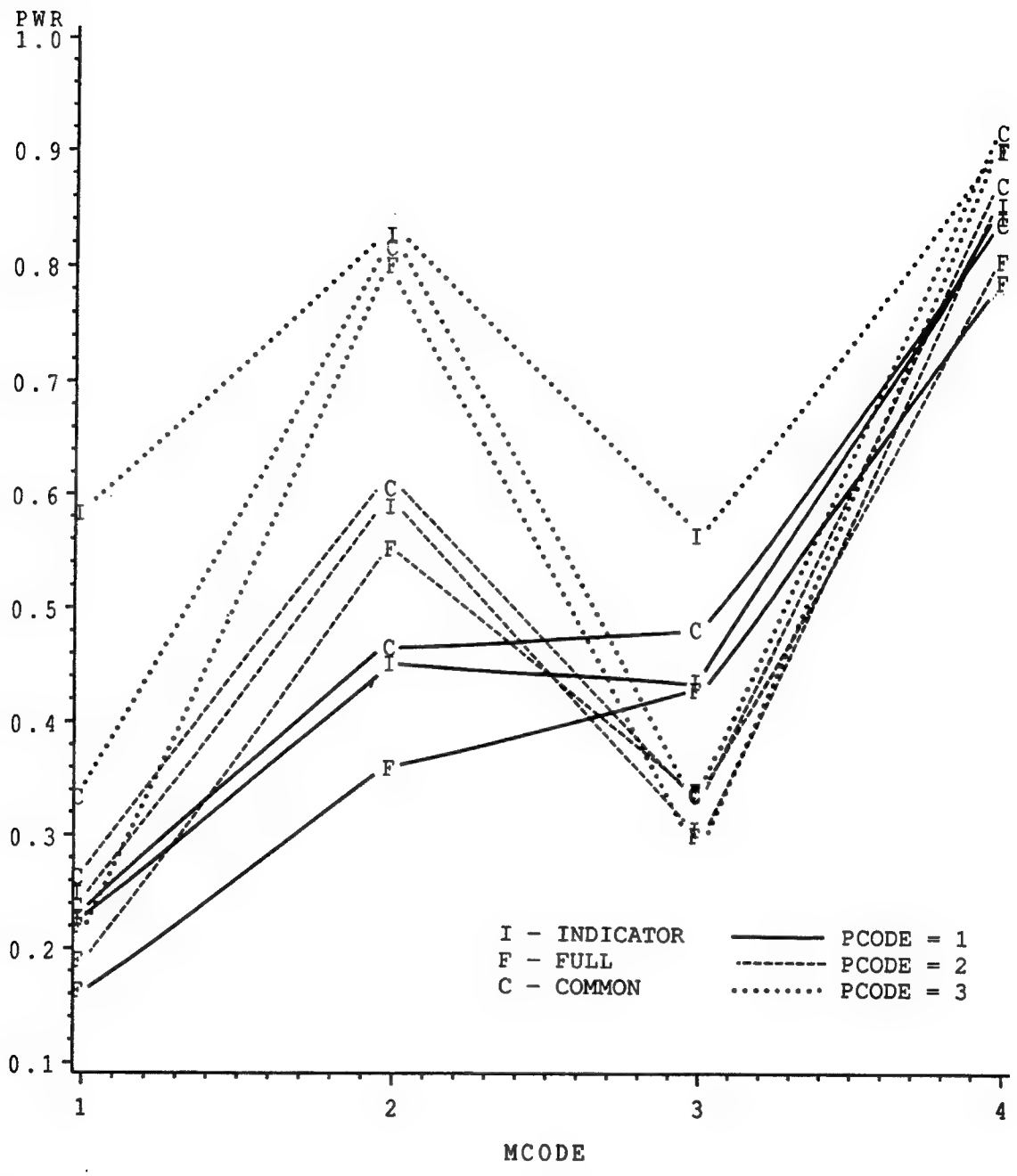
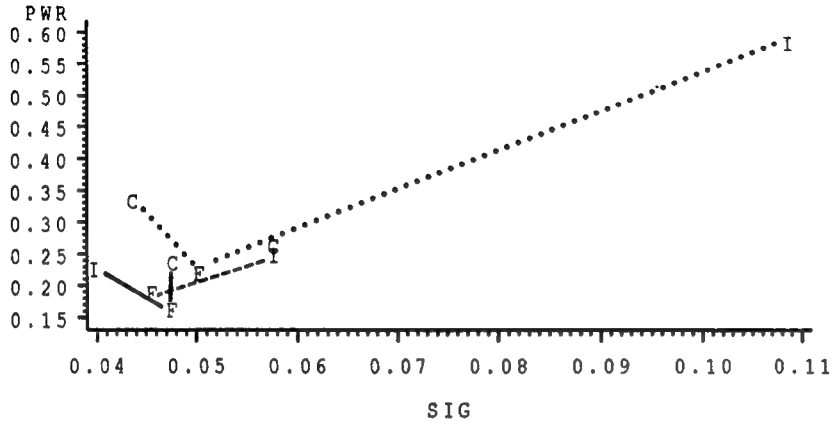
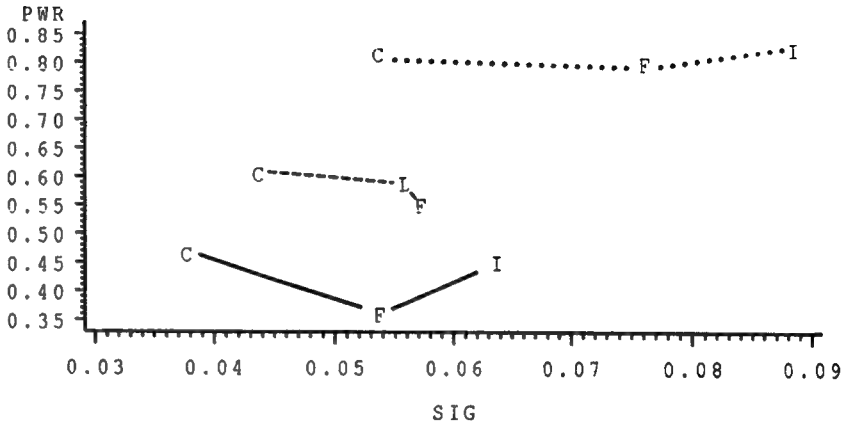


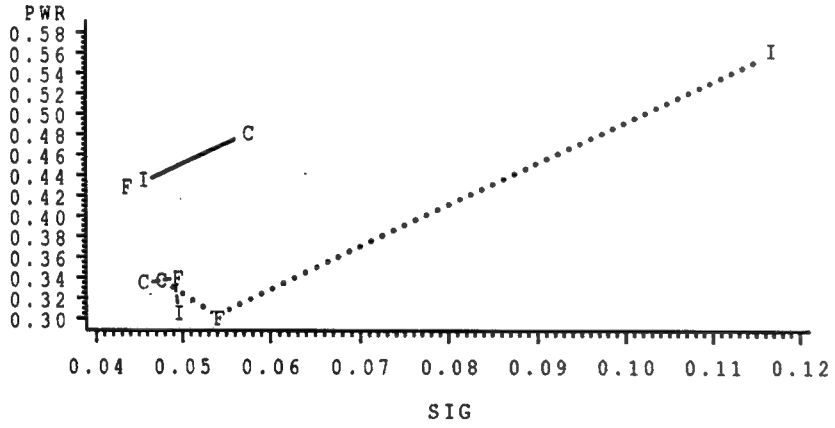
Figure 9. Power estimates when no data is missing for each of the three algorithms when data have two binary and two continuous variates, possibly dependent. For population π_1, each possible combination of the binary part occurs with probability 1/4. The conditional distribution of the continuous part is MVN($\mathbf{0}$, Σ_1), where Σ_1 is a 2x2 matrix with diagonal elements of one and off-diagonal elements of 0.5, within each multinomial cell. For population π_2, the conditional covariance matrix for the continuous part is Σ_2, where Σ_2 has ones on the diagonal and off-diagonal elements of -0.5. Several distributions for the discrete part, and several choices of mean vectors are used, as defined in Table 2.



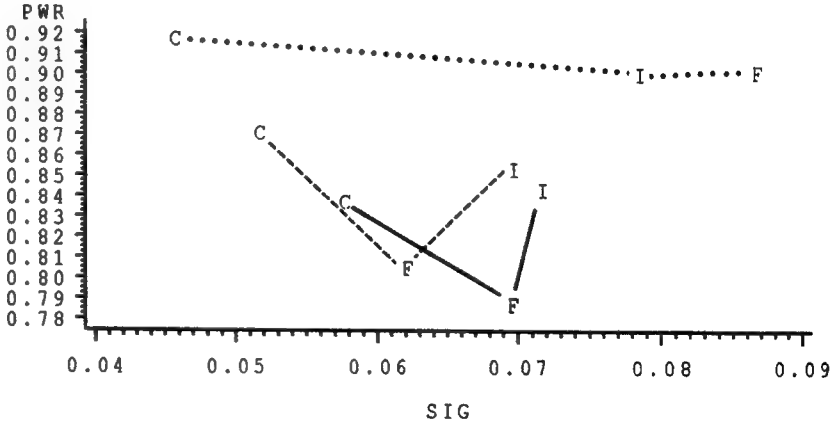
(a) MCODE = 1



(b) MCODE = 2



(c) MCODE = 3



(d) MCODE = 4

Figure 10. Plots of estimated power versus observed significance for each mean configuration in our second study when no data is missing. Symbol used is I for INDICATOR algorithm, F for FULL, and C for COMMON. The solid line corresponds to PCODE = 1, the dashed line to PCODE = 2, and the dotted line to PCODE = 3.

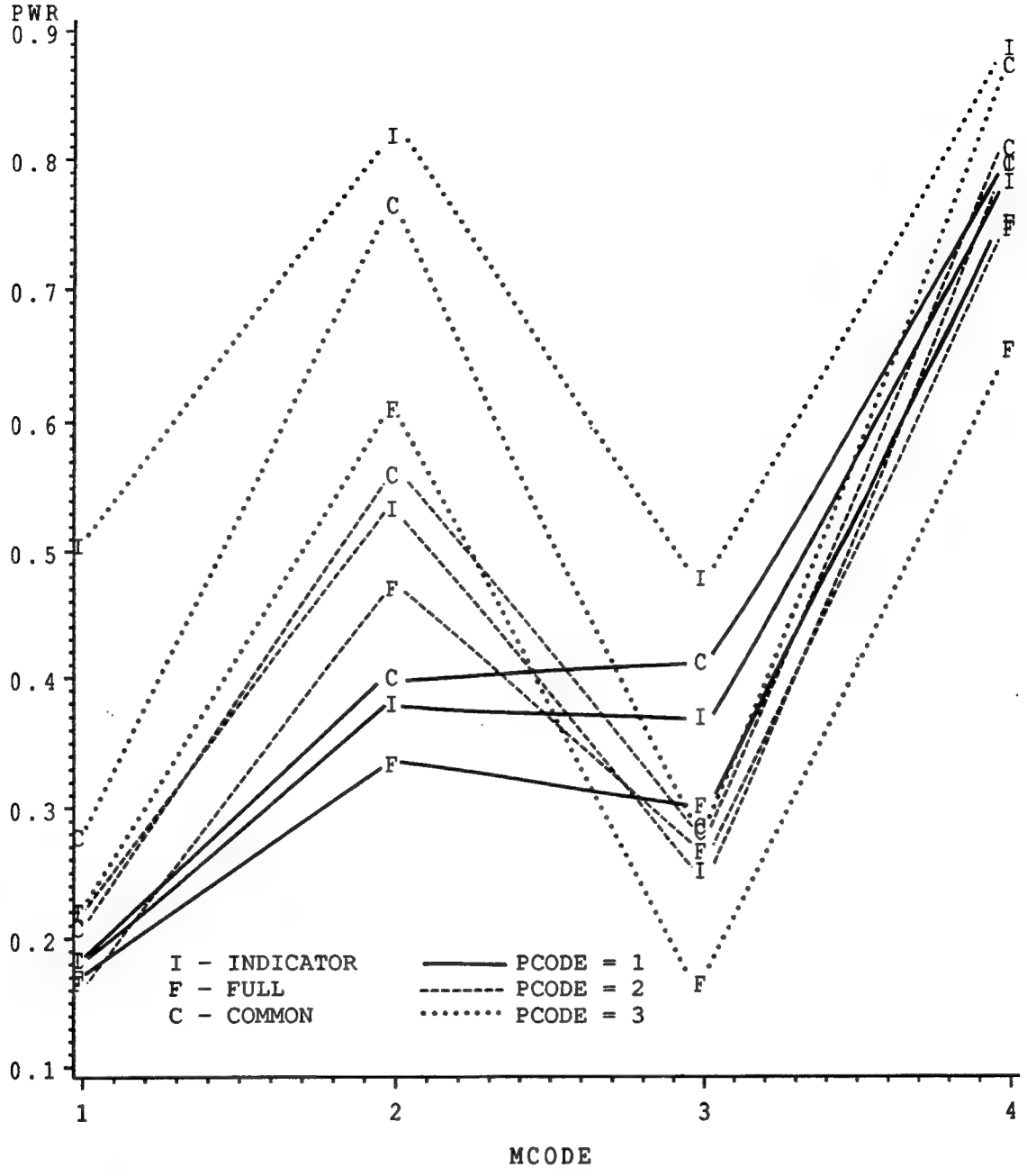
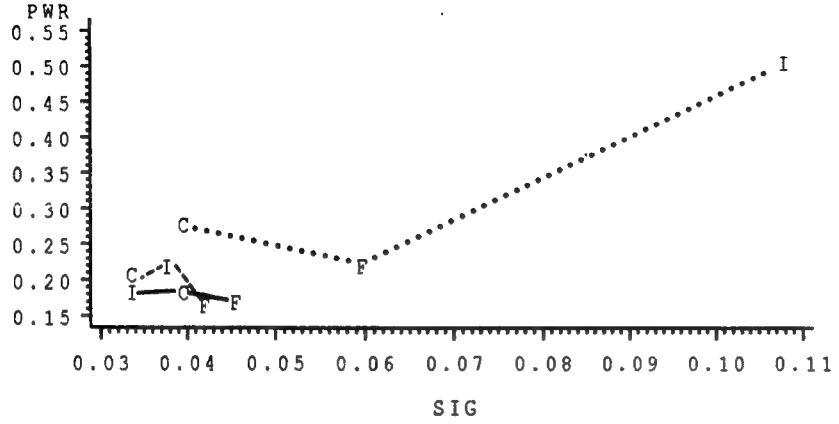
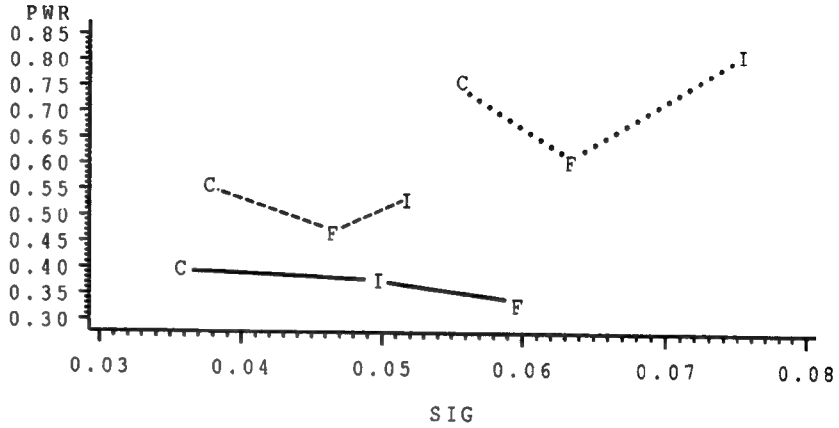


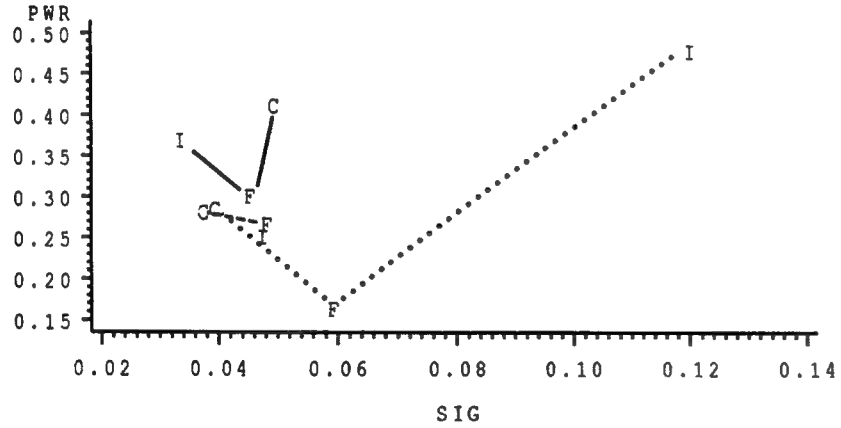
Figure 11. Power estimates when approximately 10% of the data is missing for each of the three algorithms when data have two binary and two continuous variates, possibly dependent. For population π_1, each possible combination of the binary part occurs with probability 1/4. The conditional distribution of the continuous part is MVN($\mathbf{0}$, Σ_1), where Σ_1 is a 2x2 matrix with diagonal elements of one and off-diagonal elements of 0.5, within each multinomial cell. For population π_2, the conditional covariance matrix for the continuous part is Σ_2, where Σ_2 has ones on the diagonal and off-diagonal elements of -0.5. Several distributions for the discrete part, and several choices of mean vectors are used, as defined in Table 2.



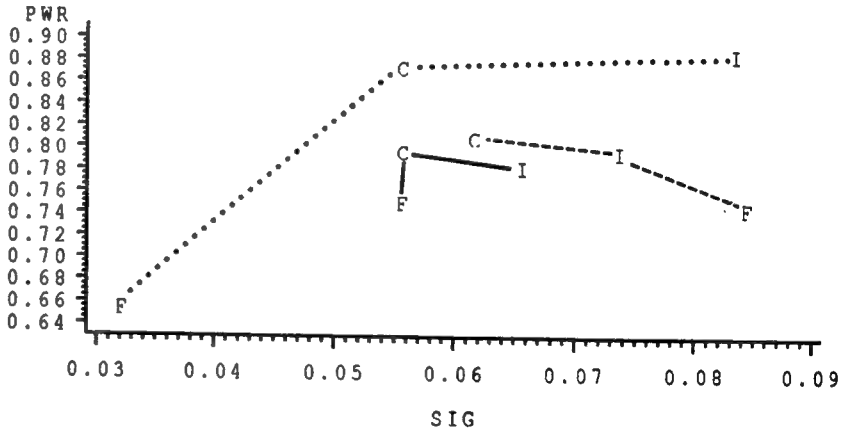
(a) MCODE = 1



(b) MCODE = 2



(c) MCODE = 3



(d) MCODE = 4

Figure 12. Plots of estimated power versus observed significance for each mean configuration in our second study when 10% of the data is missing. Symbol used is I for INDICATOR algorithm, F for FULL, and C for COMMON. The solid line corresponds to PCODE = 1, the dashed line to PCODE = 2, and the dotted line to PCODE = 3.

Prof. Thomas Ahrens
Seismological Lab, 252-21
Division of Geological & Planetary Sciences
California Institute of Technology
Pasadena, CA 91125

Prof. Keiiti Aki
Center for Earth Sciences
University of Southern California
University Park
Los Angeles, CA 90089-0741

Prof. Shelton Alexander
Geosciences Department
403 Deike Building
The Pennsylvania State University
University Park, PA 16802

Dr. Thomas C. Bache, Jr.
Science Applications Int'l Corp.
10260 Campus Point Drive
San Diego, CA 92121 (2 copies)

Prof. Muawia Barazangi
Cornell University
Institute for the Study of the Continent
3126 SNEE Hall
Ithaca, NY 14853

Dr. Douglas R. Baumgardt
ENSCO, Inc
5400 Port Royal Road
Springfield, VA 22151-2388

Dr. T.J. Bennett
S-CUBED
A Division of Maxwell Laboratories
11800 Sunrise Valley Drive, Suite 1212
Reston, VA 22091

Dr. Robert Blandford
AFTAC/TT, Center for Seismic Studies
1300 North 17th Street
Suite 1450
Arlington, VA 22209-2308

Dr. Steven Bratt
ARPA/NMRO
3701 North Fairfax Drive
Arlington, VA 22203-1714

Dale Breding
U.S. Department of Energy
Recipient, IS-20, GA-033
Office of Arms Control
Washington, DC 20585

Dr. Jerry Carter
Center for Seismic Studies
1300 North 17th Street
Suite 1450
Arlington, VA 22209-2308

Mr Robert Cockerham
Arms Control & Disarmament Agency
320 21st Street North West
Room 5741
Washington, DC 20451,

Dr. Zoltan Der
ENSCO, Inc.
5400 Port Royal Road
Springfield, VA 22151-2388

Dr. Stanley K. Dickinson
AFOSR/NM
110 Duncan Avenue
Suite B115
Bolling AFB, DC

Dr Petr Firbas
Institute of Physics of the Earth
Masaryk University Brno
Jecna 29a
612 46 Brno, Czech Republic

Dr. Mark D. Fisk
Mission Research Corporation
735 State Street
P.O. Drawer 719
Santa Barbara, CA 93102

Dr. Cliff Frolich
Institute of Geophysics
8701 North Mopac
Austin, TX 78759

Dr. Holly Given
IGPP, A-025
Scripps Institute of Oceanography
University of California, San Diego
La Jolla, CA 92093

Dr. Jeffrey W. Given
SAIC
10260 Campus Point Drive
San Diego, CA 92121

Dr. Dale Glover
Defense Intelligence Agency
ATTN: ODT-1B
Washington, DC 20301

Dan N. Hagedon
Pacific Northwest Laboratories
Battelle Boulevard
Richland, WA 99352

Dr. James Hannon
Lawrence Livermore National Laboratory
P.O. Box 808, L-205
Livermore, CA 94550

Dr. Roger Hansen
University of Colorado, JSPC
Campus Box 583
Boulder, CO 80309

Prof. David G. Harkrider
Division of Geological & Planetary Sciences
California Institute of Technology
Pasadena, CA 91125

Prof. Danny Harvey
University of Colorado, JSPC
Campus Box 583
Boulder, CO 80309

Prof. Donald V. Helmberger
Division of Geological & Planetary Sciences
California Institute of Technology
Pasadena, CA 91125

Prof. Eugene Herrin
Geophysical Laboratory
Southern Methodist University
Dallas, TX 75275

Prof. Robert B. Herrmann
Department of Earth & Atmospheric Sciences
St. Louis University
St. Louis, MO 63156

Prof. Lane R. Johnson
Seismographic Station
University of California
Berkeley, CA 94720

Prof. Thomas H. Jordan
Department of Earth, Atmospheric &
Planetary Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

Robert C. Kemerait
ENSCO, Inc.
445 Pineda Court
Melbourne, FL 32940

U.S. Dept of Energy
Max Koontz, NN-20, GA-033
Office of Research and Develop.
1000 Independence Avenue
Washington, DC 20585

Dr. Richard LaCoss
MIT Lincoln Laboratory, M-200B
P.O. Box 73
Lexington, MA 02173-0073

Prof. Charles A. Langston
Geosciences Department
403 Deike Building
The Pennsylvania State University
University Park, PA 16802

Jim Lawson, Chief Geophysicist
Oklahoma Geological Survey
Oklahoma Geophysical Observatory
P.O. Box 8
Leonard, OK 74043-0008

Prof. Thorne Lay
Institute of Tectonics
Earth Science Board
University of California, Santa Cruz
Santa Cruz, CA 95064

Dr. William Leith
U.S. Geological Survey
Mail Stop 928
Reston, VA 22092

Mr. James F. Lewkowicz
Phillips Laboratory/GPE
29 Randolph Road
Hanscom AFB, MA 01731-3010(2 copies)

Dr. Gary McCartor
Department of Physics
Southern Methodist University
Dallas, TX 75275

Prof. Thomas V. McEvilly
Seismographic Station
University of California
Berkeley, CA 94720

Dr. Keith L. McLaughlin
S-CUBED
A Division of Maxwell Laboratory
P.O. Box 1620
La Jolla, CA 92038-1620

Prof. Bernard Minster
IGPP, A-025
Scripps Institute of Oceanography
University of California, San Diego
La Jolla, CA 92093

Prof. Brian J. Mitchell
Department of Earth & Atmospheric Sciences
St. Louis University
St. Louis, MO 63156

Mr. Jack Murphy
S-CUBED
A Division of Maxwell Laboratory
11800 Sunrise Valley Drive, Suite 1212
Reston, VA 22091 (2 Copies)

Dr. Keith K. Nakanishi
Lawrence Livermore National Laboratory
L-025
P.O. Box 808
Livermore, CA 94550

Prof. John A. Orcutt
IGPP, A-025
Scripps Institute of Oceanography
University of California, San Diego
La Jolla, CA 92093

Dr. Howard Patton
Lawrence Livermore National Laboratory
L-025
P.O. Box 808
Livermore, CA 94550

Dr. Frank Pilotte
HQ AFTAC/TT
1030 South Highway A1A
Patrick AFB, FL 32925-3002

Dr. Jay J. Pulli
Radix Systems, Inc.
201 Perry Parkway
Gaithersburg, MD 20877

Prof. Paul G. Richards
Lamont-Doherty Earth Observatory
of Columbia University
Palisades, NY 10964

Mr. Wilmer Rivers
Multimax Inc.
1441 McCormick Drive
Landover, MD 20785

Dr. Alan S. Ryall, Jr.
Lawrence Livermore National Laboratory
L-025
P.O. Box 808
Livermore, CA 94550

Dr.Chandan K. Saikia
Woodward Clyde- Consultants
566 El Dorado Street
Pasadena, CA 91101

Mr. Dogan Seber
Cornell University
Inst. for the Study of the Continent
3130 SNEE Hall
Ithaca, NY 14853-1504

Secretary of the Air Force
(SAFRD)
Washington, DC 20330

Office of the Secretary of Defense
DDR&E
Washington, DC 20330

Thomas J. Sereno, Jr.
Science Application Int'l Corp.
10260 Campus Point Drive
San Diego, CA 92121

Dr. Michael Shore
Defense Nuclear Agency/SPSS
6801 Telegraph Road
Alexandria, VA 22310

Prof. David G. Simpson
IRIS, Inc.
1616 North Fort Myer Drive
Suite 1050
Arlington, VA 22209

Dr. Jeffrey Stevens
S-CUBED
A Division of Maxwell Laboratory
P.O. Box 1620
La Jolla, CA 92038-1620

Prof. Brian Stump
Los Alamos National Laboratory
EES-3
Mail Stop C-335
Los Alamos, NM 87545

Prof. Tuncay Taymaz
Istanbul Technical University
Dept. of Geophysical Engineering
Mining Faculty
Maslak-80626, Istanbul Turkey

Prof. M. Nafi Toksoz
Earth Resources Lab
Massachusetts Institute of Technology
42 Carleton Street
Cambridge, MA 02142

Dr. Larry Turnbull
CIA-OSWR/NED
Washington, DC 20505

Dr. Karl Veith
EG&G
5211 Auth Road
Suite 240
Suitland, MD 20746

Prof. Terry C. Wallace
Department of Geosciences
Building #77
University of Arizona
Tuscon, AZ 85721

Dr. William Wortman
Mission Research Corporation
8560 Cinderbed Road
Suite 700
Newington, VA 22122

ARPA, OASB/Library
3701 North Fairfax Drive
Arlington, VA 22203-1714

HQ DNA
ATTN: Technical Library
Washington, DC 20305

Defense Technical Information Center
Cameron Station
Alexandria, VA 22314 (2 Copies)

TACTEC
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201 (Final Report)

Phillips Laboratory
ATTN: GPE
29 Randolph Road
Hanscom AFB, MA 01731-3010

Phillips Laboratory
ATTN: TSML
5 Wright Street
Hanscom AFB, MA 01731-3004

Phillips Laboratory
ATTN: PL/SUL
3550 Aberdeen Ave SE
Kirtland, NM 87117-5776 (2 copies)

Dr. Michel Campillo
Observatoire de Grenoble
I.R.I.G.M.-B.P. 53
38041 Grenoble, FRANCE

Dr. Kin Yip Chun
Geophysics Division
Physics Department
University of Toronto
Ontario, CANADA

Prof. Hans-Peter Harjes
Institute for Geophysic
Ruhr University/Bochum
P.O. Box 102148
4630 Bochum 1, GERMANY

Prof. Eystein Husebye
NTNF/NORSAR
P.O. Box 51
N-2007 Kjeller, NORWAY

David Jepsen
Acting Head, Nuclear Monitoring Section
Bureau of Mineral Resources
Geology and Geophysics
G.P.O. Box 378, Canberra, AUSTRALIA

Ms. Eva Johannisson
Senior Research Officer
FOA
S-172 90 Sundbyberg, SWEDEN

Dr. Peter Marshall
Procurement Executive
Ministry of Defense
Blacknest, Brimpton
Reading FG7-FRS, UNITED KINGDOM

Dr. Bernard Massinon, Dr. Pierre Mechler
Societe Radiomana
27 rue Claude Bernard
75005 Paris, FRANCE (2 Copies)

Dr. Svein Mykkeltveit
NTNT/NORSAR
P.O. Box 51
N-2007 Kjeller, NORWAY (3 Copies)

Dr. Jorg Schlittenhardt
Federal Institute for Geosciences & Nat'l Res.
Postfach 510153
D-30631 Hannover , GERMANY

Dr. Johannes Schweitzer
Institute of Geophysics
Ruhr University/Bochum
P.O. Box 1102148
4360 Bochum 1, GERMANY

Trust & Verify
VERTIC
Carrara House
20 Embankment Place
London WC2N 6NN, ENGLAND